

UNIVERSITÉ AIX-MARSEILLE I – Université de Provence

U.F.R. L.A.C.S.

N° attribué par la bibliothèque

/ _ / _ / _ / _ / _ / _ / _ /

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ AIX-MARSEILLE I

Formation doctorale : Cognition, Langage, Education

présentée par

Liang MA

le 28 avril 2008

**LA COARTICULATION EN FRANÇAIS ET EN CHINOIS :
ÉTUDE EXPÉRIMENTALE ET MODÉLISATION**

**Sous la direction de M. le Professeur Noël Nguyen et
M. le Professeur Pascal Perrier**

JURY

Alain Marchal, Directeur de recherche, président du jury

Phil Hoole, Professeur, rapporteur

Rudolph Sock, Professeur, rapporteur

Daniel Hirst, Directeur de recherche

Michel Grenié, Maître de conférences

Noël Nguyen, Professeur, co-directeur

Pascal Perrier, Professeur, co-directeur

Remerciements

Je remercie tout d'abord Noël Nguyen et Pascal Perrier d'avoir accepté de diriger ce travail. Je remercie Noël Nguyen de m'avoir accueillie au Laboratoire Parole et Langage et pour ses conseils avisés. Je remercie également Pascal Perrier pour toutes les discussions et conseils précieux.

Merci à Daniel Hirst pour les bons conseils et les discussions, merci de m'avoir prêté ses livres en phonétique chinoise.

Merci à Christine Meunier et Cécile Fougeron pour leurs discussions et leurs recommandations.

Merci à Robert Espesser et Cyril Deniaud pour leurs aides techniques et à Simone Dufour, Isabelle Vincent-Marlien et Armelle Bonpain pour leur gentillesse.

Merci à tous mes camarades thésards du labo qui m'ont toujours aidée durant ces années...surtout à Céline De Looze, Vincent Aubanel, Amina Chentir qui ont apporté de nombreuses corrections de français à la rédaction de cette thèse.

Merci surtout aux membres du jury d'avoir accepté de juger ce travail et d'avoir été présents. Pour finir, je remercie ma famille et mon mari pour leurs soutiens et leurs encouragements durant tout cette période.

TABLE DES MATIÈRES

INTRODUCTION.....	6
PREMIERE PARTIE : FONDEMENTS THEORIQUES.....	8
CHAPITRE 1 : LA COARTICULATION, QUELQUES GENERALITES	8
1.1 Les idées principes des modèles de la coarticulation	9
1.1.1 La prédominance syllabique	9
1.1.2 La différenciation entre base vocalique et consonne (Öhman, 1966,1967)	10
1.1.3 La propagation de traits : Modèle “look-ahead” (Henke, 1966)	12
1.1.4 La coproduction.....	13
1.1.5 Optimalité (Keating, 1988)	14
1.2 Exploitation de ces idées principes pour l’analyse de la variabilité des mouvements et des sons de la parole	15
1.2.1 Les modèles inspirés par les propositions d’Öhman	15
1.2.2 Interprétation basée sur l’hypothèse de coproduction	16
1.2.2.1 Chevauchement gestuel.....	16
1.2.2.2 Cohésion gestuelle	17
1.2.2.3 Timing intrinsèque	18
1.2.3 Planification de coarticulation.....	19
1.2.3.1 Planification fondée sur des critères phonologiques	20
1.2.3.2 Planification fondée sur des critères moteurs	21
1.3 Les modèles généraux de contrôle de production de la parole.....	23
1.3.1 Le “carrier model” (Dang <i>et al.</i> 2005).....	23
1.3.2 “Task-dynamics” et “Coordinative structure” : Le modèle des laboratoires Haskins	26
1.3.3 Planification optimale	28
1.3.3.1 Degrés de liberté en excès et réseaux séquentiels.....	28
1.3.3.2 Le modèle DIVA : l’optimalité dans un cadre syllabique.....	31
CHAPITRE 2 : DESCRIPTION PHONOLOGIQUE SIMPLIFIEE ET COARTICULATION POUR LE CHINOIS ET POUR LE FRANÇAIS.....	33
2.1 Description phonologique simplifiée pour le chinois et pour le français	34
2.1.1 L’inventaire vocalique	34
2.1.1.1 L’inventaire vocalique en chinois	35
2.1.1.2 L’inventaire vocalique en français	35
2.1.2 Le rôle de la syllabe.....	36
2.1.2.1 La syllabe en chinois	37
2.1.2.2 La syllabe en français.....	40
a. “Resyllabification”	40
b. Le rôle de la syllabe pour la segmentation en français.....	41
c. Le rôle de la syllabe pour la prosodie du français.....	42
2.2 La coarticulation en français et en chinois	44
2.2.1 La coarticulation en français	44
2.2.1.1 La coarticulation d’arrondissement.....	44
2.2.1.2 La coarticulation de nasalité en français	45
2.2.1.3 La coarticulation de voyelle à voyelle.....	46
2.2.2 La coarticulation en chinois	48
2.2.2.1 La coarticulation entre les syllabes	49
2.2.2.2 La coarticulation dans la syllabe.....	50
2.2.2.3 La coarticulation de N ₁ C ₂ en chinois	50
CHAPITRE 3 : BILAN : ENJEUX ET DEMARCHE METHODOLOGIQUE	52
3.1 Les enjeux de notre modélisation de planification	52
3.1.1 Planifier dans quel espace ? Avec quelle complexité ?.....	52

3.1.2 Le modèle interne et la planification	54
3.1.3 Critère de planification	57
3.1.4 Interaction linguistique de planification	58
3.2 Méthodologie	59
3.2.1 Une étude expérimentale en français et en chinois.....	59
3.2.2 Une modélisation du contrôle moteur.....	60
3.2.2.1 La méthodologie d'une modélisation du contrôle moteur.....	60
3.2.2.2 Hypothèse de planification et d'exécution du mouvement.....	61
A. Hypothèse de planification	61
a. Hypothèse de planification de voisinage	61
b. Hypothèse de planification syllabique.....	61
c. Hypothèse de planification inspirée du modèle d'Öhman (1967)	62
B. Hypothèse d'exécution du mouvement	62
a Hypothèse d'exécution séquentielle	62
b Hypothèse d'exécution selon Öhman.....	62
DEUXIEME PARTIE : METHODOLOGIE	63
CHAPITRE 4 : DEMARCHE POUR NOTRE MODELISATION DU CONTROLE MOTEUR DE LA	
PAROLE	63
4.1 Modèle biomécanique	63
4.2 Constitution de la base de simulation	65
4.2.1 Génération des formes de langue par la Méthode de Monte-Carlo.....	66
4.2.2 Génération des patrons formantiques (F ₁ , F ₂ , F ₃).....	66
4.3 Modèle interne statique	71
4.3.1 Introduction d'apprentissage du modèle interne statique.....	71
4.3.2 Les fonctions radiales de bases	72
4.3.3 Principe de l'apprentissage	74
4.4 Inversion : des commandes phonémiques aux commandes motrices	76
4.4.1 Corrélats acoustico-perceptifs des phonèmes	77
4.4.2 Inversion du modèle interne par optimisation d'un critère	79
4.4.3 Exemples d'inversion	79
CHAPITRE 5 : METHODOLOGIE EXPERIMENTALE	83
5.1 Recueil et description des données	83
5.1.1 Recueil des données.....	83
5.1.2 Description du corpus	84
5.1.3 Les sujets.....	85
5.2 L'étiquetage	85
5.2.1 Le logiciel pour l'étiquetage	85
5.2.2 L'étiquetage acoustique	86
5.2.3 Modification automatique de l'étiquetage	87
5.3 Enregistrement et modification du contour du palais	89
TROISIEME PARTIE : RESULTATS	92
CHAPITRE 6 : ANALYSE DES DONNEES ARTICULATOIRES	92
6.1 Anticipation non adjacente : Effet de la voyelle V₂ sur la voyelle V₁	92
6.1.1 Position atteinte par la langue pour V ₁ =/a/ dans les séquences /ata/, /ati/ et /atu/	93
6.1.2 Position atteinte par la langue pour V ₁ =/i/ dans les séquences /ita/, /iti/ et /itu/.....	98
6.1.3 Position atteinte par la langue pour V ₁ =/u/ dans les séquences /uta/, /uti/ et /utu/	101
Résumé :	104
6.1.4 Position atteinte par la langue pour V ₁ =/a/ dans les séquences /aka/ et /aku/	105
6.1.5 Position atteinte par la langue pour V ₁ =/i/ dans les séquences /ika/ et /iku/	108
6.1.6 Position atteinte par la langue pour V ₁ =/u/ dans les séquences /uka/ et /uku/.....	111
Résumé :	111

6.2 Anticipation adjacente : Effet de la voyelle V₂ sur C	114
6.2.1 Position atteinte par la langue pour C=/t/ dans les séquences /ata/, /ati/ et /atu/	114
6.2.2 Position atteinte par la langue pour C=/t/ dans les séquences /ita/, /iti/ et /itu/	118
6.2.3 Position atteinte par la langue pour C=/t/ dans les séquences /uta/, /uti/ et /utu/	121
Résumé :	124
6.2.4 Position atteinte par la langue pour C=/k/ dans les séquences /aka/ et /aku/	124
6.2.5 Position atteinte par la langue pour C=/k/ dans les séquences /ika/ et /iku/	128
6.2.6 Position atteinte par la langue pour C=/k/ dans les séquences /uka/ et /uku/	131
Résumé :	134
Bilan générale :	134
6.3 Analyse temporelle	135
6.3.1 Analyse de la durée entre la consonne C et la cible voyelle V ₂	137
6.3.1.1 Séquences /V ₁ tV ₂ /	137
6.3.1.2 Séquences /V ₁ kV ₂ /	138
6.3.2 Analyse de la durée entre la cible voyelle V ₁ et la cible voyelle V ₂	140
6.3.2.1 Séquences /V ₁ tV ₂ /	140
6.3.2.2 Séquences /V ₁ kV ₂ /	141
Résumé :	142
CHAPITRE 7 : SIMULATIONS DE SEQUENCES VCV	143
7.1 Hypothèse de la planification	144
7.1.1 Hypothèse de planification de voisinage	144
7.1.2 Hypothèse de planification syllabique	146
7.1.3 Hypothèse de planification inspirée du modèle d'Öhman	147
7.2 Simulation des séquences VCV	150
7.2.1 L'exécution séquentielle du mouvement	151
7.2.1.1 Les simulations pour une planification de type voisinage avec exécution séquentielle	152
7.2.1.2 Les simulations pour une planification de type syllabique avec exécution séquentielle	153
7.2.1.3 Les simulations sans planification avec exécution séquentielle	154
7.2.2 L'exécution du mouvement selon Öhman	156
7.2.2.1 Les simulations pour une planification selon Öhman avec exécution selon Öhman	158
7.2.2.2 Les simulations pour une planification de type voisinage avec exécution selon Öhman	159
7.2.2.3 Effet de décalage de l'initiation du geste consonantique	160
7.3 Evaluation systématique des simulations dans le plan sagittal	162
7.3.1 Hypothèse de planification de type voisinage avec exécution séquentielle	163
a. Effet de la voyelle V ₂ sur la consonne /k/	163
b. Effet de la voyelle V ₂ sur la voyelle V ₁	163
7.3.2 Hypothèse de planification syllabique avec exécution séquentielle	169
a. Effet de la voyelle V ₂ sur la consonne /k/	169
b. Effet de la voyelle V ₂ sur la voyelle V ₁	169
7.3.3 Hypothèse de planification inspirée du modèle d'Öhman avec exécution selon Öhman	173
a. Effet de la voyelle V ₂ sur la consonne /k/	173
b. Effet de la voyelle V ₂ sur la voyelle V ₁	173
7.3.4 Deux autres modèles s'appuyant sur des hypothèses de planification de voisinage avec exécution selon Öhman	179
Résumé :	184
CHAPITRE 8 : DISCUSSION	185
CONCLUSION	196
BIBLIOGRAPHIE	198

Introduction

La coarticulation est un phénomène qui traduit le fait que les caractéristiques de la production de la parole associées à un phonème donné varient de manière importante en fonction des phonèmes adjacents. Elle est la conséquence de deux facteurs principaux : (1) la stratégie de planification régissant les gestes articulatoires de la parole, et (2) les caractéristiques du système physique de production mis en jeu lors de l'exécution des gestes de la parole. Comme d'autres mouvements humains, les mouvements de la parole sont planifiés selon des stratégies optimales de commandes motrices, en minimisant les « efforts » du locuteur dans le système moteur, tout en atteignant les objets de la tâche. Cependant, sur le niveau supérieur de ces mécanismes généraux de contrôle, la planification de la parole pourrait être contrainte par les facteurs phonologiques de la langue, qui pourraient influencer la longueur des séquences sur lesquelles les principes d'optimisation s'appliquent.

Nous avons pour objectif d'étudier l'influence linguistique potentielle sur la planification de la coarticulation dans des séquences de type VCV. Pour cela, nous proposons un corpus en français et en chinois pour une analyse expérimentale. Le français et le chinois ont été considérés comme ayant un statut de la syllabe différent dans leur langue. Pour le chinois, la syllabe est l'unité la plus fondamentale de la structure phonologique, tandis que la force de la syllabe semble être moins forte pour le français. Notre but est de parvenir à différencier les propriétés spécifiques de stratégie liées à la langue dans la coarticulation d'anticipation. Nous cherchons également à tester quantitativement différentes hypothèses de planification et d'exécution du mouvement pour les séquences VCV en exploitant une modélisation du contrôle moteur, dont le comportement sera comparé aux données expérimentales.

Dans un premier temps, des données articulatoires ont été recueillies pour trois sujets français et trois sujets chinois au moyen d'un système électromagnétique (EMMA). Le corpus est composé de 15 phrases VCV où les voyelles utilisées étaient /a i u/ et les consonnes étaient la vélaire /k/ et l'alvéolaire /t/. Nous avons effectué des étiquetages acoustiques et des étiquetages articulatoires sur les données. La coarticulation d'anticipation de voyelle V_2 sur la voyelle V_1 et de voyelle V_2 sur la consonne C, dans les séquences de V_1CV_2 , a été analysée sur les positionnements des articulateurs de la langue (pour quatre capteurs placés sur la langue, à partir de l'apex, notés T_1 , T_2 , T_3 et T_4). Des analyses ANOVA (Mesures Répétées) et des tests de post-hoc ont été effectués pour chaque sujet. Nous avons ensuite comparé les comportements d'anticipation des locuteurs dans les deux groupes de langues.

Dans un deuxième temps, nous avons élaboré un modèle de contrôle optimal de la planification que nous avons appliqué à un modèle biomécanique de la langue (Payan et Perrier, 1997, Perrier, 2003). La modélisation du contrôle moteur passe par une première phase consistant à élaborer un modèle direct qui décrit des relations entre les commandes motrices et les caractéristiques spectrales du signal de la parole, que nous qualifierons de “modèle interne”. Ce modèle sera ensuite exploité, dans une seconde phase, afin d’exploiter les synergies et les antagonismes musculaires lors de la planification gestuelle d’une séquence de la parole. Cette phase consiste à inverser les commandes motrices associées à la génération des caractéristiques spectrales cibles souhaitées. Mais la relation entre les commandes motrices et les caractéristiques spectrales n’est pas biunivoque, car pour une cible donnée il existe plusieurs commandes motrices qui lui sont associées. Pour faire une telle inversion, nous avons choisi d’optimiser certains critères sur l’ensemble de la séquence qui intègrent des contraintes orientées vers le locuteur et des contraintes orientées vers l’auditeur. Nous avons proposé trois hypothèses de planification : (1) l’hypothèse de voisinage : une planification globale pour les séquences V_1CV_2 . (2) l’hypothèse syllabique qui donne un statut spécifique à la syllabe CV_2 . (3) l’hypothèse inspirée du modèle d’Öhman qui donne un statut spécifique à la transition vocalique V_1V_2 . Une fois les cibles planifiées, il est intéressant de savoir comment générer les mouvements entre les cibles planifiées. Nous proposons de tester également deux modèles d’exécution du mouvement, l’exécution séquentielle et l’exécution selon Öhman.

Notre thèse est composée de trois parties sub-divisées en huit chapitres. Dans une première partie, nous proposons les fondements théoriques de notre travail, composé d’une présentation générale des théories de coarticulation (dans le premier chapitre), d’une description phonologique simplifiée en français et en chinois (dans le deuxième chapitre), et les enjeux et les démarches méthodologiques (dans le troisième chapitre). La deuxième partie est consacrée à la méthodologie de modélisation du contrôle optimal (dans le quatrième chapitre) et à la méthodologie expérimentale avec la présentation du corpus, de l’étiquetage, et du post traitement des données articulatoires (dans le cinquième chapitre). La troisième partie présente les résultats de l’analyse des données articulatoires (dans le sixième chapitre), les résultats des simulations en exploitant la modélisation de contrôle moteur. Différentes hypothèses sont proposées (dans le septième chapitre) et une discussion (dans le huitième chapitre) est menée pour la synthèse des ces résultats. En conclusion, nous proposons des perspectives d’étude dans le cadre de la coarticulation.

Première partie : Fondements théoriques

Chapitre 1 : La coarticulation, quelques généralités

Introduction

La coarticulation montre la complexité de la relation entre le concept de segments phonologiques et sa manifestation dans la production de la parole. La parole met en jeu un processus qui prend en entrée des segments discrets et dont la sortie est un continuum articulatoire et acoustique. La question que nous posons ici est de savoir comment les segments phonologiques s'influencent les uns les autres dans leur réalisation.

La production de la parole passe par le contrôle de tâches motrices permettant l'émission de l'onde sonore, vecteur physique du message linguistique à destination des articulateurs. Selon FOWLER *et al.* (1980), il existerait deux niveaux de contrôle en production de la parole : le premier niveau sélectionne des entités abstraites correspondant aux unités phonologiques. Il constitue le niveau supérieur, responsable de tout ce qui est structuration des objectifs liés à la chaîne phonologique. Un deuxième niveau exécute la tâche motrice pour réaliser ces entités. C'est le niveau inférieur, responsable de tout ce qui est articulation. Ces propositions peuvent être mises en regard de celles de Whalen (1990) qui considère que la parole est planifiée. La planification de la parole correspondrait alors au niveau supérieur de contrôle. Dans cette étape de planification, une représentation sous forme d'unités discrètes constitue l'entrée du mécanisme du système de production de la parole. Le deuxième niveau serait celui de l'exécution des tâches planifiées par le système moteur.

Dans cette perspective, la coarticulation serait la conséquence de deux facteurs principaux : (1) une stratégie de planification au niveau supérieur, (Whalen, 1990; Vatikiotis-Bateson *et al.* 1994; Dang *et al.* 2006) et (2) les caractéristiques du système physique de production mis en jeu lors de l'exécution des gestes de la parole, tels que le couplage mécanique intrinsèque entre articulateurs ou propriétés dynamiques des articulateurs (raideur et inertie) (Browman et Goldstein, 1989; Lindblom, 1963; Perrier *et al.* 1996). Les caractéristiques articulatoires et acoustiques associées à la production d'un phonème donné peuvent alors varier de manière importante en fonction des phonèmes environnants.

Traditionnellement on considère en phonétique deux types de coarticulation (1) la coarticulation dite "de gauche à droite" ou "carryover" en anglais qui rend compte de l'influence des phonèmes précédents, (2) la coarticulation "de droite à gauche" ou anticipation ou coarticulation progressive qui traduit l'influence des phonèmes suivants. L'anticipation et

le carryover sont considérés comme les résultats de processus différents. Le carryover serait plutôt vu comme une conséquence de la physique du système de production, le point de départ pour chaque phonème étant les positions atteintes par les articulateurs dans les phonèmes précédents. Dans ce cas, la configuration articuloire atteinte pour le deuxième phonème est différente selon la configuration articuloire atteinte lors du premier phonème. Par contre, l'anticipation a lieu seulement si le locuteur peut prévoir et anticiper les phonèmes à venir. Elle est donc vue comme le résultat de la stratégie de planification au niveau supérieur.

Ainsi, une bonne connaissance de l'anticipation dans une langue donnée aide à comprendre le contrôle moteur de la parole dans cette langue. Le but de notre étude est de parvenir à différencier les propriétés spécifiques des stratégies de planification liées à une langue donnée dans la coarticulation. Pour cela, dans cette thèse, nous nous concentrerons sur la mesure des effets de l'anticipation. Dans ce chapitre, quelques points théoriques relatifs à la coarticulation sont abordés dans la première partie à travers la présentation des données caractéristiques de la littérature ; ensuite des exploitations de ces points théoriques de la coarticulation, et enfin quelques modèles de contrôle dans la production de la parole seront décrits.

1.1 Les idées princeps des modèles de la coarticulation

1.1.1 La prédominance syllabique

Kozhevnikov et Chistovich (1965) observent que l'articulation de la voyelle dans une syllabe CV commence dès le début de celle-ci si elle ne requiert pas pour sa réalisation de mouvements contradictoires avec ceux nécessaires pour la production de la consonne. L'exécution de la consonne initiale et celle de la voyelle d'une syllabe CV seraient ainsi initiées simultanément par le locuteur. La syllabe est ainsi considérée par ces deux auteurs comme l'unité de base de l'articulation.

“All the movements of a vowel which are not contradictory to the articulation of the consonant begin with the beginning of the syllable” (P122).

Leurs expériences sur le geste d'arrondissement labial en russe pour la voyelle arrondie leur ont permis de caractériser le domaine d'extension de la coarticulation. Ce résultat est cohérent avec l'hypothèse de la syllabe articuloire. Les auteurs observent en effet que la coarticulation ne s'étend pas au-delà des frontières de la syllabe CV.

Wood (1991) a confirmé l'hypothèse syllabique de Kozhevnikov et Chistovich dans son interprétation de la variabilité des mouvements de la langue. Il a étudié les données cinéradiographiques d'un locuteur suédois prononçant les séquences /'ɛbe/ /'ʃi:se/ /'ç:sar/ et /i'su:daʃu/. Les mesures montrent que pour chaque syllabe ce locuteur initie le geste vers la consonne dans la dernière moitié du segment précédent. Quand il n'y a pas d'antagonisme entre les gestes de la consonne et de la voyelle dans une syllabe, la voyelle de la nouvelle syllabe commence aussi dans cette portion de la séquence. Ainsi, la voyelle /e/ est lancée avec la consonne /b/ simultanément. Si la langue est contrainte différemment pour la voyelle et la consonne, le locuteur retarde le début du geste vers la voyelle. Ainsi, dans la syllabe /ʃi:/ de /'ʃi:se/, l'initiation du geste de la langue pour la voyelle /i:/ commence un peu plus tard que celui de la consonne initiale, parce que la langue est encore occupée avec le geste palatovélaire de la consonne /ʃ/. Ces observations confirment l'hypothèse selon laquelle la syllabe serait un élément fondamental d'organisation des stratégies de la coarticulation anticipatoire : les gestes vers la consonne ne commencent que juste avant la réalisation de cette consonne et, dans certains cas, la voyelle suivante est initiée au même moment dans la parole.

1.1.2 La différenciation entre base vocalique et consonne (Öhman, 1966,1967)

Öhman (1966,1967) a étudié la coarticulation dans les séquences V_1CV_2 pour le suédois et l'anglais, dans un corpus où trois consonnes occlusives voisées /b d g/ ont été utilisées. Il a montré que les transitions des formants de V_1C dépendent de la voyelle suivante V_2 . De façon similaire, les transitions des formants de CV_2 sont influencées par la voyelle précédente V_1 . Il a ainsi trouvé que les effets de coarticulation peuvent se faire sentir au-delà des limites de la syllabe CV. Un modèle de coarticulation pour la séquence VCV a été proposé à partir de ces observations. Dans ce modèle, l'effet de la coarticulation dans la séquence VCV est interprété comme la conséquence d'un statut particulier des transitions de voyelle à voyelle, les consonnes étant considérées alors comme des perturbations, localisées dans le temps, de cette base vocalique.

“The data suggest a physiological model in terms of which the VCV articulations are represented by a basic diphthongal gesture with an independent stop-consonant gesture superimposed on its transitional portion.” (P151).

Selon l'hypothèse d'Öhman, les systèmes articulatoires mis en jeu dans les productions des voyelles devraient se distinguer de ceux des consonnes. Öhman a alors noté que la forme du conduit vocal n'est pas pertinente dans sa totalité pendant la tenue d'une occlusive. Le geste articulatoire de la voyelle peut ainsi être exécuté par la langue pendant la production de la consonne. Cette observation l'a amené à suggérer de dissocier dans la langue l'activité de trois ensembles de muscles, qui auraient des représentations séparées dans le Système Nerveux Central du locuteur. Les commandes articulatoires pourraient être alors transmises par trois canaux indépendamment les uns des autres. Ainsi trois régions séparées de la langue pourraient être indépendamment contrôlées : le corps de la langue (utilisé pour la production des voyelles), une région de l'apex de la langue (utilisée pour la production des consonnes alvéolaires), et la région dorsale (utilisée pour la production des consonnes vélaires).

“...the production of vowel-stop-vowel utterance of certain languages seemed to involve two simultaneous gestures, a diphthongal gesture of tongue body articulator and a superimposed constrictory gesture of the apical or dorsal articulators.” (1966, p310)

Perkell (1969) confirme l'hypothèse d'Öhman d'un point de vue biomécanique et physiologique sur la base des analyses des images cinéradiographiques. Sur cette base expérimentale, il a proposé un modèle physiologique dans lequel il sépare les actions générées pour la production de la voyelle et celles qui sont liées à la production des consonnes. La plus grande partie du conduit vocal est affectée à la fois par la production des voyelles et celle des consonnes, mais en général les mêmes organes semblent se comporter différemment sous l'influence des deux différentes classes. Les articulations des consonnes par la langue et les lèvres sont généralement plus rapides et plus complexes que les articulations des voyelles. Perkell a constaté aussi que, dans une certaine mesure, il y a une division anatomique. Par exemple, l'apex de la langue est plus impliqué dans l'articulation des consonnes, tandis que le corps de la langue est en activité dans les articulateurs des consonnes et des voyelles. Les différences générales dans la vitesse, la complexité, la précision du mouvement suggèrent que différents types de muscles pourraient être responsables pour la production des consonnes et des voyelles. Il semblerait que l'articulation des voyelles soit accomplie principalement par la grande, et plus lente, musculature extrinsèque de la langue. Les consonnes utiliseraient une musculature intrinsèque, plus courte et plus rapide. La coarticulation semble être le résultat de l'interaction de deux systèmes neuromusculaires spécifiques.

“It is probable that articulation of vowels is accomplished principally by the large, slower extrinsic tongue musculature which controls tongue position. On the other hand, consonant articulation requires the addition of the precise, more complex, and faster function of the smaller, intrinsic tongue musculature.” (P61).

1.1.3 La propagation de traits : Modèle “look-ahead” (Henke, 1966)

Cette hypothèse repose d’abord sur une conception purement phonologique de la production de la parole inspirée par les travaux de Chomsky et Halle (1968) selon laquelle la production de la parole consisterait à réaliser pour chaque son élémentaire un certain nombre de traits caractéristiques. On peut donc concevoir au niveau phonologique que la planification d’une séquence de parole consiste à définir une succession de vecteurs de traits pour chaque phonème. Mais la spécification d’un phonème ne fait pas systématiquement appel à tous les traits. Il y a en général un certain nombre de traits qui est aussi “neutre” ou “sous spécifié”. Ceci introduit un certain nombre de degré de liberté dans la planification phonologique de la séquence, et c’est cela que l’hypothèse de propagation de traits (*“feature-spreading”*) et le modèle “look-ahead” (Henke, 1966) exploitent.

L’hypothèse de propagation de traits, lorsqu’elle traite d’un segment dont un des traits spécifiques est neutre dans un ou plusieurs des segments précédents, considère que la spécification de ce trait se rétropropage aux segments précédents jusqu’à ce que cela devienne contraire à la spécification phonologique intrinsèque d’un de ces segments. Par exemple, dans la séquence /istu/, où le trait considéré est l’arrondissement des lèvres, la première marque d’arrondissement est observée dès la fin de la voyelle /i/. Les consonnes /s/ et /t/ sont en effet neutres pour le trait de l’arrondissement, tandis que la voyelle /i/ exige des lèvres écartées (non arrondies).

Ceci est formalisé dans le modèle “look-ahead” dont les variables d’entrée sont les phonèmes spécifiés par un ensemble de traits. Chaque trait porte une valeur “+”, “-” ou “non spécifié”. Quand l’entrée est “non spécifié”, le modèle rétropropage la prochaine valeur spécifiée (+ ou -) de ce trait et ceci se termine lorsque la prochaine valeur qui suit est en contradiction avec une entrée “+” ou “-”. Il est clair que selon le modèle “look-ahead”, la coarticulation dépasse les limites de la syllabe CV.

1.1.4 La coproduction

La coproduction est une théorie proposée par un groupe de chercheurs des Laboratoires Haskins : Fowler (1977, 1980), Kelso *et al.* (1986) et Saltzman (1989). L'idée sous-jacente à la notion de coproduction est celle selon laquelle les mouvements articulatoires de la parole seraient le résultat de la superposition des gestes, chacun des gestes étant étroitement associé à la production d'un phonème donné. Le geste est défini comme un membre d'une famille de mouvements articulatoires fonctionnellement équivalents qui sont activement contrôlés pour un certain but de parole.

"The term gesture is used..... to denote a member of a family of functionally equivalent articulatory movement patterns that are actively controlled with reference to a given speech relevant goal " (Saltzman et Munhall, 1989)

La théorie de coproduction propose alors que les gestes se chevauchent l'un l'autre dans la parole (Bell-Berti et Harris, 1981 ; Fowler, 1977, 1980 ; Harris, 1984 ; Hardcastle, 1981 ; Saltzman et Munhall, 1989). Ce chevauchement des gestes est considéré comme la source du phénomène de coarticulation. La coarticulation traduit le fait que l'influence des gestes liés à plusieurs segments adjacents peut généralement être discernée dans des mesures acoustiques ou articulatoires. L'effet de ce chevauchement gestuel est bien illustré par la figure 1.1 (Löfqvist, 1990). La figure 1.1 montre une représentation de deux gestes avec différents degrés de chevauchement liés à la variation de débit. Lorsque le débit est lent, deux gestes séparés sont observés. Lorsque le débit est rapide, un seul geste est observé et, à un débit intermédiaire, la trajectoire articulatoire montre les traces de deux gestes individuels. On voit que selon cette théorie, les gestes qui forment un segment conservent une stabilité interne quelque soit le débit. Ils ont une dimension temporelle intrinsèque.

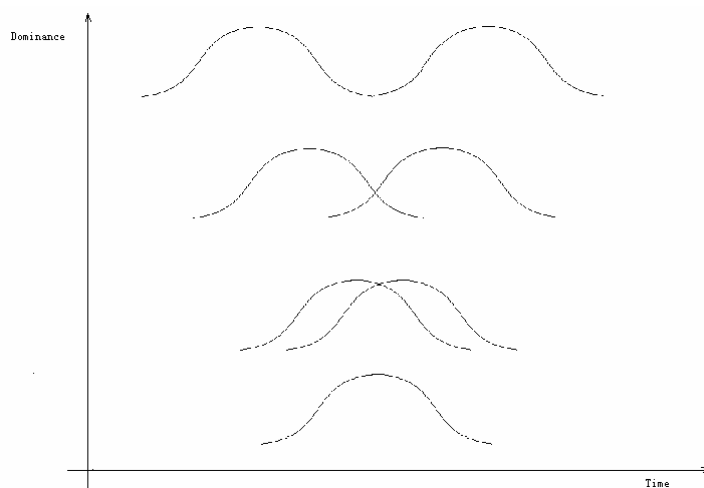


Figure 1.1 représentation de deux gestes avec différents degrés de chevauchement (Löfqvist, 1990)

1.1.5 Optimalité (Keating, 1988)

Keating (1988b) a proposé une optimisation gestuelle en minimisant la distance parcourue dans l'espace articulatoire dans son modèle de fenêtre. La variabilité joue un rôle important dans ce modèle. Keating a proposé que, pour une dimension articulatoire, chaque valeur de trait d'un segment soit associée avec une série de valeurs spatiales possibles, i.e. les valeurs observées doivent se trouver entre la valeur minimale et maximale. Cette série de valeurs est appelée une fenêtre, donc un modèle de fenêtre (*"window model"*) est un modèle spatial. Une fenêtre représente la variabilité contextuelle de valeur d'un trait. La fenêtre étroite reflète peu de variation contextuelle et la fenêtre large reflète une grande variation contextuelle. D'après Keating, la largeur d'une fenêtre est relative à la spécification du trait. La figure 1.2 montre le principe du modèle de fenêtre. Dans cette figure, la fenêtre étroite représente des segments spécifiés associés avec certain trait, et la fenêtre large représente le segment non spécifié (ou peu spécifié) pour ce trait. Le panneau à gauche montre une trajectoire articulatoire entre 3 segments spécifiés. Le panneau à droite montre un cas où il y a un segment non spécifié représenté par une fenêtre large au milieu de deux segments spécifiés. Il est clair qu'en cherchant une distance minimale, le modèle de fenêtre permet une trajectoire articulatoire plus courte pour le deuxième cas que pour le premier.

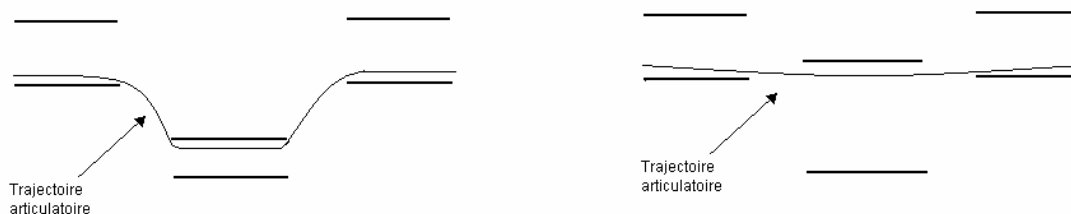


Figure 1.2: modèle de fenêtre: optimisation des trajectoires dans l'espace articulatoire
(Keating 1988b)

Le modèle de fenêtre donne la possibilité d'assigner un segment approprié (ou spécifié) à une cible, donc une fenêtre étroite, dans l'espace articulatoire. En ce qui concerne le segment non spécifié, Keating a proposé dans la théorie de sous-spécification que les segments phonétiquement non spécifiés ne seraient pas caractérisés par des cibles spécifiques dans le plan contrôle moteur.

"When phonetic rules build trajectories between segments, an unspecified segment will contribute nothing of its own to the trajectory". (Keating 1988a, 281)

1.2 Exploitation de ces idées princeps pour l'analyse de la variabilité des mouvements et des sons de la parole

1.2.1 Les modèles inspirés par les propositions d'Öhman

Recasens a exploité l'hypothèse d'Öhman (1966) selon laquelle la consonne et la voyelle impliqueraient des gestes différents pour interpréter et modéliser des données sur la coarticulation VCV. Les productions des voyelles nécessiteraient le contrôle articulaire sur la position du corps de la langue dans son ensemble, tandis que les consonnes impliqueraient un contrôle articulaire dépendant du lieu de constriction ou l'occlusion du conduit vocal. Recasens (1984, 1987, 2002) a étudié l'effet de la coarticulation de la voyelle sur la consonne (V-C) et de la voyelle sur la voyelle (V-V) dans des données articulaires et acoustiques, en catalan et en espagnol pour les séquences VCV. Les résultats montrent différentes influences de V_2 sur C et de V_2 sur V_1 selon la nature de la consonne C. Il a alors interprété ces données dans le cadre proposé par Öhman en suggérant que les gestes fortement contraints pourraient limiter l'étendue spatiale et temporelle de l'effet de la coarticulation. Il a ainsi proposé le concept de degré de contrainte articulaire (DAC) pour quantifier ce phénomène de résistance à la coarticulation. L'influence de la voyelle V_2 sur la consonne C dans une séquence V_1CV_2 est d'autant plus faible que le DAC de la consonne est grand. Plus encore, la coarticulation V_1-V_2 décroît elle aussi quand le DAC de la consonne intermédiaire croît. L'anticipation de la voyelle V_2 sur la voyelle V_1 est ainsi observée, mais elle dépend de la contrainte articulaire apportée par la production de la consonne intervocalique.

Carré (1991,1995) a modélisé les séquences VCV à l'aide d'un modèle DRM (Distinctive Region Model). Il a pour cela repris l'idée introduite par Öhman, d'une superposition de la consonne sur la base du geste vocalique. Un corpus de séquences V_1-V_2 en français a été enregistré dans un premier temps pour établir les trajectoires V_1-V_2 dans l'espace formantique. Carré (1995) a ensuite synthétisé ces trajectoires V_1-V_2 à partir des commandes gestuelles proposées par le modèle DRM (Voir Boë et Perrier, 1990 pour une critique de ces commandes). Une fois ces transitions V_1-V_2 modélisées, la modélisation d'une séquence V_1CV_2 peut se réaliser par la superposition de consonnes intervocaliques à cette base vocalique. Selon Carré, les résultats de cette modélisation sont cohérents avec les données expérimentales d'Öhman (1966).

1.2.2 Interprétation basée sur l'hypothèse de coproduction

1.2.2.1 Chevauchement gestuel

Le chevauchement des gestes a été observé pour le geste de la langue utilisant l'électropalatographie (EPG), (Hardcastle, 1985 ; Marchal, 1988). Marchal (1988) a examiné les séquences de $V_1C_1C_2V_2$ en français dans lesquelles les consonnes étaient deux consonnes linguales successives, comme /kt/ ou /kl/, et les voyelles étaient /a/, /i/ et /u/. L'organisation temporelle des gestes articulatoires des consonnes a été analysée. Le résultat montre que la stratégie employée par le locuteur pour produire les consonnes successives C_1C_2 est en faveur d'une hypothèse de coproduction. Donc, selon Marchal, l'organisation des gestes articulatoires ne peut pas être interprété comme la simple concaténation de segments assimilés. La coarticulation est plutôt une conséquence de chevauchement entre les gestes articulatoires.

Boyce (Boyce *et al.* 1990) a aussi expliqué ses données sur la trajectoire labiale dans des séquences /i C₁....C_n u/ par l'hypothèse du chevauchement entre gestes. La figure 1.3 présente deux gestes théoriques associés avec deux segments en chevauchement, et une trajectoire lissée. Boyce a suggéré que le deuxième geste est un geste de la protrusion pour une voyelle arrondie /u/ et le premier geste est celui de la consonne précédente. Le point auquel le geste de protrusion commence varie selon la taille et la forme du geste de la consonne, et le timing relatif de ces 2 gestes. La figure 1.4 présente les mouvements de protrusion de la lèvre supérieure, pour six paires de séquences /kiktluk/ et /kiktlik/ enregistrées par un locuteur. Pour les séquences /kiktluk/, la protrusion de la voyelle /u/ montre une forme de "deux phases". Il y a typiquement deux maxima d'accélération entre le début de protrusion et un maximum de protrusion. Le premier apparaît juste après le début du mouvement. Le maximum principal de protrusion apparaît environ 100 ms après le début acoustique de la voyelle /u/. Une inflexion secondaire entre le début de protrusion et le maximum principal correspond au début de la deuxième phase. Le deuxième maximum d'accélération du mouvement de protrusion apparaît en ce moment. Après avoir comparé les séquences /kiktlik/ avec les séquences /kiktluk/, Boyce a indiqué que la première phase observée pour les séquences /kiktluk/ est due aux consonnes devant la voyelle /u/ plutôt qu'au résultat entier du geste de la voyelle. Cette inflexion secondaire pourrait refléter la frontière d'interaction entre les gestes pour les segments adjacents.

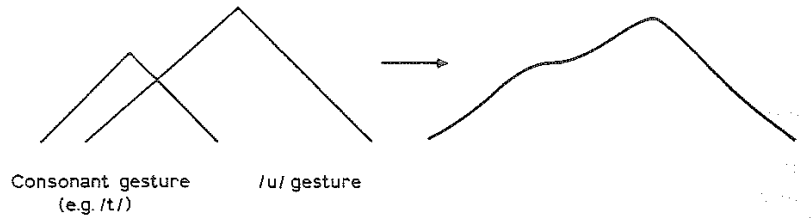


Figure 1.3 : Le chevauchement théorique des deux gestes (à gauche) et une trajectoire lissée du chevauchement (à droite). (Boyce *et al.* 1990).

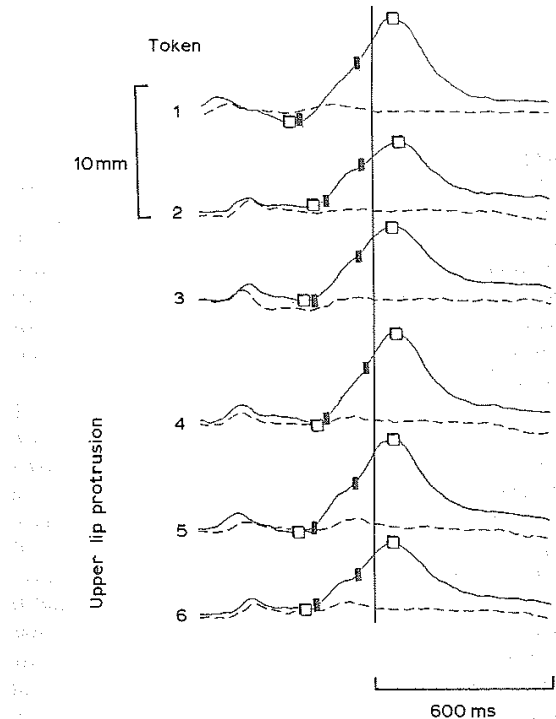


Figure 1.4 : Mouvement de protrusion de la lèvre supérieure, pour six paires de séquences /kiktluk/ vs /kiktlik/. Le trait vertical représente le début acoustique de la deuxième voyelle. Le trait solide représente la séquence /kiktluk/ et le trait pointillé pour la séquence /kiktlik/ ; le symbole de “□” représente le début et la fin de protrusion ; le symbole de “■” représente le maximum d’accélération. (Boyce *et al.* 1990).

1.2.2.2 Cohésion gestuelle

Certains phénomènes de coarticulation ont été expliqués par des caractéristiques de cohésion particulières entre certains gestes (Saltzman et Munhall, 1989 ; Browman et Goldstein, 1986, 1989). Browman et Goldstein (1986, 1989) ont ainsi proposé, dans leur théorie de la Phonologie Articulatoire, que les primitives phonologiques sont des constellations gestuelles (“*cohesive bundles*”). La cohésion des gestes est conçue en terme de stabilité de la phase ou de relations spatiotemporelles entre les gestes élémentaires dans une constellation donnée. Selon ces auteurs, ce sont ces primitives phonologiques qui pourraient être observées directement à partir du mouvement articulatoire. Dans la théorie de la Phonologie Articulatoire de Browman et Goldstein, les unités phonologiques sont définies

comme des unités spécifiques de l'action articulatoire. Les gestes sont organisés par une propriété à la fois temporelle et spatiale. L'importance de cette théorie est l'identification des unités phonologiques avec des gestes et la proposition de constellations gestuelles.

L'intérêt de cette hypothèse a été bien illustré par des expériences de perturbation labiale et de leur impact sur les mouvements glottiques. Ainsi lors de la production des obstruents sourdes (Munhall, Löfqvist et Kelso, 1986), des compensations laryngées ont été observées quand la lèvre inférieure a été perturbée pendant la production de l'obstruente. Particulièrement, si la lèvre inférieure est tirée en bas de façon inattendue juste avant la fermeture bilabiale, le geste laryngé est retardé. Ceci explique l'hypothèse de cohésion entre les gestes labiaux et glottiques proposée par Löfqvist et Yoshioka (1981).

1.2.2.3 Timing intrinsèque

Au delà de la variabilité du positionnement des articulateurs qui peuvent expliquer les concepts de superposition et de cohésion gestuelle introduits par la théorie de la production, la notion de gestes a aussi offert un cadre pour interpréter les patrons temporels d'une séquence de parole. En effet, dans la conception des laboratoires Haskins, le geste est indissociable de la notion de système dynamique. La formalisation sous forme de modèle de la théorie de coproduction a en effet été réalisée dans le cadre du modèle de la dynamique de la tâche ("*Task dynamics*") (Saltzman et Munhall, 1989 ; Kelso *et al.* 1986) où le geste est décrit comme le mouvement vers un attracteur dynamique dans l'espace des variables du conduit vocal qui caractérisent la géométrie du conduit vocal visée pour l'unité phonologique en cours de réalisation. Ainsi les caractéristiques temporelles du geste sont celles de l'attracteur dynamique. Elles sont intrinsèques au système dynamique. Fowler (1980) considère ainsi que la coproduction de geste pourrait être vue comme le couplage dynamique des systèmes dynamiques caractérisant chaque geste. Les évolutions temporelles des articulateurs concernés par ces gestes seraient alors inhérentes aux caractéristiques dynamiques des systèmes en interaction. Fowler (1980) parle de timing intrinsèque, c'est à dire découlant de la spécification des gestes, par opposition à un timing extrinsèque où l'évolution temporelle des articulateurs serait spécifiée par le Système Nerveux Central.

Cette proposition a été affirmée par Saltzman et Munhall (1989) qui ont proposé que le timing effectif soit la combinaison des effets des caractéristiques dynamiques des gestes et d'un séquençage externe déterminant les instanciations de chaque geste ou chaque groupe de gestes. Mais ils conservent cette idée clé selon laquelle le déroulement temporel des

articulateurs est en partie déterminé par les caractéristiques des gestes coproduits.

Bell-Berti *et al.* (1979, 1981, 1991) ont exploité ce concept pour expliquer des données expérimentales sur l'anticipation de l'arrondissement labial et de l'abaissement du velum. Ces auteurs ont en effet observé que dans une séquence V_1CV_2 , où V_2 est arrondie (ou nasale) alors que V_1 et C ne le sont pas, l'arrondissement (ou l'abaissement du velum) commence dans la consonne à moment fixe par rapport au début acoustique de V_2 . Cela donne l'idée principale du modèle "time locked" (Bell-Berti et Harris 1981). Contrairement au modèle "look-ahead", la durée d'anticipation dans le modèle "time locked" est indépendante de la longueur de la chaîne phonétique précédente. Dans le cadre de la théorie de geste dynamique, la régularité temporelle observée serait le résultat non pas d'un contrôle extérieur, mais bien celui du timing intrinsèque du geste d'arrondissement (ou d'abaissement du velum).

1.2.3 Planification de coarticulation

Ces derniers travaux sur l'anticipation de geste de protrusion labial ou d'abaissement vélaire renvoient à une question plus générale : l'anticipation gestuelle en parole n'est le fait que d'une prise en compte de caractéristiques temporelles intrinsèques comme le suggéraient Bell-Berti et ses collègues ou bien un jeu de phénomènes plus complexes de planification par le Système Nerveux Central. Selon Whalen (1990), les phénomènes de coarticulation résulteraient bien prioritairement de stratégies de planification. Il considère en effet que bien que quelques aspects de coarticulation soient attribuables à la dynamique, la plupart d'entre eux semblent être la conséquence de variations de commandes centrales.

Pour étayer ses propositions, Whalen (1990) a mis au point une expérience où les locuteurs ont été forcés de prononcer une partie de phrase dépourvue de sens avant que la phrase entière ne soit connue. Les éléments inconnus sont une voyelle ou une consonne. Dans la grande majorité des cas, il observe une anticipation des segments à venir connus, lors qu'aucune variabilité ne semble pouvoir être associée à ces mêmes segments s'ils sont inconnus. C'est donc bien un acte intentionnel et non l'interaction physique entre articulateurs ou gestes qui induisent la variabilité anticipatoire. Par contre, l'effet de carryover est présent dans les deux cas : avec un segment inconnu et un segment connu. Whalen confirme ainsi l'hypothèse classique qui dit que le carryover est différent de l'anticipation et il suggère que l'effet de carryover pourrait être plutôt d'origine physique que planifié.

Et en effet, de nombreuses données recueillies au cours des 30 dernières années soutiennent l'hypothèse d'une coarticulation planifiée : Henke (1966), Benguérel et Cowan

(1974), Perkell (1990), Abry et Lallouache (1995) sur l'anticipation de protrusion des lèvres, Wood (1994) sur l'anticipation du mouvement de la langue, et Hamlet et Stone (1981) sur l'anticipation du mouvement de mâchoire. Des travaux plus récents de modélisation intégrant un modèle physique de la langue montrent aussi que, dans des conditions normales de débit, aucune anticipation ne semble pouvoir être imputée au système physique seul (Perrier *et al*, 2004). Mais tous les travaux expérimentaux qui soutiennent l'hypothèse de la planification n'ont pas été interprétés de manière unique : certains y voient en effet la conséquence de stratégies purement linguistiques alors que d'autres y voient la conséquence de stratégies d'optimisation motrice.

1.2.3.1 Planification fondée sur des critères phonologiques

Benguerel et Cowan (1974) ont exploité le modèle “look-ahead” afin d'étudier la coarticulation de protrusion de la lèvre supérieure en français. Le geste de protrusion de la voyelle non arrondie précédée d'une séquence consonantique de longue durée est observé. Les résultats montrent que dans la séquence $/V_u C_1 C_2 \dots C_n V_r/$ (V_u = voyelle non arrondie, V_r = voyelle arrondie), le geste de protrusion de la lèvre supérieure de la voyelle arrondie $/V_r/$ commence à partir de la première consonne $/C_1/$, si aucune des consonnes C_i n'est spécifiée pour le trait d'arrondissement.

Perkell (1990) a lui aussi exploité les hypothèses émises par Henke (1966), mais il les a associées à des contraintes d'ordre physiques telles que celles qu'ont mentionnées Bell-Berti et ses collègues. Ceci a donné naissance à un modèle hybride de l'anticipation. Le modèle hybride prédit que conformément au modèle “look-ahead” l'anticipation du mouvement articulaire commencerait aussitôt que la spécification phonologique l'autorise, mais il impose une contrainte sur la variation temporelle de ce trait lorsque l'on s'approche des phonèmes pour lequel le trait est spécifié : cette dernière phrase de l'anticipation serait essentiellement déterminée par les caractéristiques dynamiques des articulateurs mis en jeu.

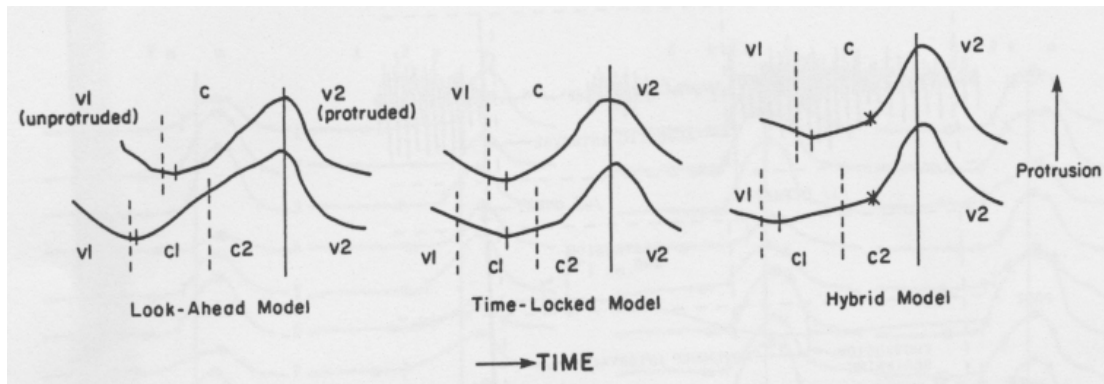


Figure 1.5 Représentation des modèles “look-ahead”, “time locked” et “hybrid” pour la protrusion des lèvres (Perkell, 1990). La ligne verticale continue représente le début acoustique de la voyelle arrondie V2. Le symbole “|” sur les courbes représente le début du mouvement. Le symbole “*” représente le point du maximum d’accélération du mouvement. Les traits pointillés représentent des frontières acoustiques entre les segments phonétiques.

Dans la figure 1.5, les deux premiers panneaux présentent le modèle “look-ahead” qui anticipe le mouvement de protrusion le plus tôt possible, et le modèle “time locked” qui prédit une anticipation à un moment fixe par rapport au début acoustique de la voyelle arrondie. Le dernier panneau présente le modèle hybride où le geste de protrusion des lèvres a deux composantes, une phase initiale et une phase seconde plus rapide. Les deux phases sont séparées l’une de l’autre par un point à une position maximum d’accélération du mouvement (Le symbole “*”). La phase initiale commence juste après la première voyelle non arrondie V₁ suivie par plusieurs consonnes comme dans le modèle “look-ahead”, mais le début de la seconde phase est toujours fixé par rapport au début de la voyelle arrondie V₂ indépendamment du nombre de consonnes comme dans le modèle “time locked”.

En conclusion, d’après Perkell, les effets de la coarticulation sont le résultat des influences conjuguées d’une planification fondée sur des critères linguistiques de la prise en compte de contraintes.

1.2.3.2 Planification fondée sur des critères moteurs

Abry et Lallouache (1995, 1996) ont proposé un Modèle d’Expansion du Mouvement (MEM) qui donne la possibilité d’observer des stratégies d’anticipation variables en fonction du locuteur. Abry et Lallouache rejoignent, pour le français, le modèle hybride de Perkell et essaient d’expliquer la variabilité en comparant le modèle “look-ahead” et le modèle “time locked” pour plusieurs sujets.

Pour étudier l’anticipation à travers des chaînes de consonnes longues, la phrase est constituée de groupes consonantiques à la jointure des mots. La phrase « Ces deux Sixtes

sculptèrent » a été utilisée. Les résultats montrent que le mouvement de protrusion dans une séquence /i C₁...C_n y/ commence pour tous les locuteurs de plus en plus tôt en fonction de l'augmentation du nombre de consonnes intervocaliques par rapport à la voyelle /y/.

La durée du cluster consonantique est variable selon les locuteurs. Cette utilisation est caractérisée par un coefficient d'expansion, dont la signification est donnée dans la figure 1.6. L'axe horizontal représente l'intervalle acoustique commençant à la fin de première voyelle non arrondie et finissant du début de la deuxième voyelle arrondie. L'axe vertical est la durée du mouvement, défini du début du mouvement au début acoustique de la voyelle arrondie. La durée du mouvement de la transition simple /iy/, dans le cas de deux voyelles voisines, est d'environ 140ms. L'intervalle acoustique est autour de 100 ms.

Le Modèle d'Expansion du Mouvement (MEM) propose que, pour des séquences de plusieurs consonnes intervocaliques, la durée du mouvement augmente linéairement avec l'intervalle acoustique, et ceci selon une stratégie propre à chaque locuteur. Le coefficient d'expansion du mouvement correspond à la pente de cette augmentation linéaire, et à la pente 1 coefficient égal à 1. Il varie selon les locuteurs. La durée d'anticipation peut aussi, pour certain locuteur, atteindre celle que prédirait le modèle "look-ahead" (par exemple pente de 0.93 pour un locuteur), mais ce n'est pas systématiquement le cas pour tous les locuteurs. Le comportement des locuteurs peut être très différent à la fois pour le modèle "look-ahead" et pour le modèle "time locked" (par exemple, la pente de 0.54 et 0.42 pour deux autres locuteurs). Cette durée d'anticipation ne peut cependant pas être inférieure à un seuil minimum (140 ms) qui correspond notamment à la durée du mouvement mesurée quand deux voyelles se succèdent immédiatement. La durée du mouvement est fortement expansible, mais relativement peu compressible.

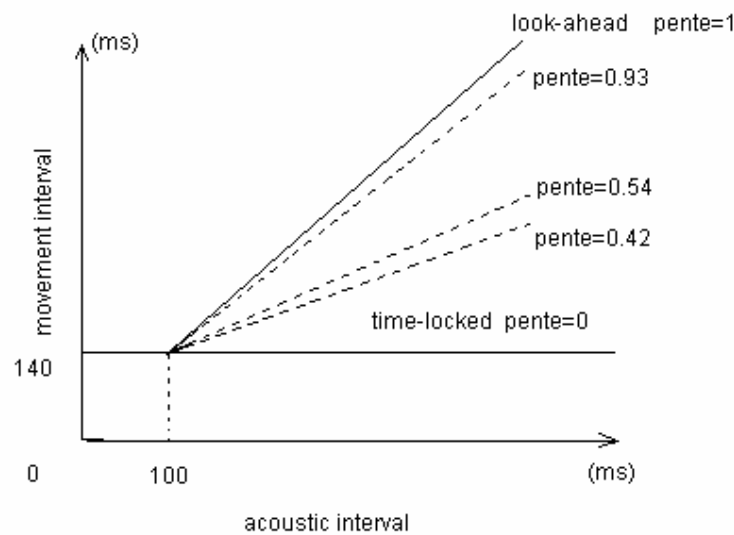


Figure 1.6 Le modèle MEM (Abry et Lallouache, 1996)
Paramétré sur la pente ou le coefficient d'expansion du mouvement.

Le modèle MEM rejette l'hypothèse de l'interaction dynamique entre les gestes. Pour Abry et Lallouache, les données relevées sur les locuteurs sont uniquement le résultat d'une planification centrale. Ils montrent les stratégies des locuteurs : lorsque la durée du groupe consonantique augmente, la durée de l'anticipation augmente selon une relation linéaire. Le modèle "look-ahead" semble être une stratégie extrême de l'anticipation, mais la réalité est différente parce que locuteur utilise selon Abry et Lallouache la stratégie d'une façon qui lui est propre. Cependant, le modèle prend aussi les contraintes des propriétés physiques du système périphérique. La durée minimale du mouvement pour les deux voyelles successives a été prévue dans le modèle en prenant les contraintes des propriétés biomécaniques des articulateurs impliqués dans le mouvement.

1.3 Les modèles généraux de contrôle de production de la parole

Nous allons présenter dans cette section quelques modèles de production de la parole sur le contrôle moteur. Nous aborderons les principes généraux de la modélisation et du contrôle du mouvement.

1.3.1 Le "carrier model" (Dang *et al.* 2005)

Dang *et al.* (2005) ont proposé un modèle de génération des commandes motrices pour contrôler un modèle biomécanique du conduit vocal qui intègre au niveau de la planification un phénomène de coarticulation : le "carrier model", inspiré très largement des hypothèses

d'Öhman (1966) sur la dissociation du canal vocalique et du canal consonantique.

La figure 1.7 donne une représentation synoptique du modèle de la production de la parole proposé par Dang *et al.*. L'entrée du modèle est un ensemble de configurations articulaires "type" caractéristiques des phonèmes de la langue. A chacun des phonèmes est associée une configuration type unique. Ces sont les objectifs discrets que la production de la parole cherche à atteindre. Dang *et al.* (2005) les appellent les cibles phonétiques typiques. Le "carrier model" gère une évolution des cibles articulaires d'une cible typique à la suivante en prenant en compte les influences respectives des éléments d'une séquence les uns sur les autres. Les sorties du "carrier model" sont les cibles planifiées. Ces cibles planifiées sont alors employées afin de commander les articulateurs du modèle biomécanique (DANG, 2004) pour produire des mouvements articulaires.

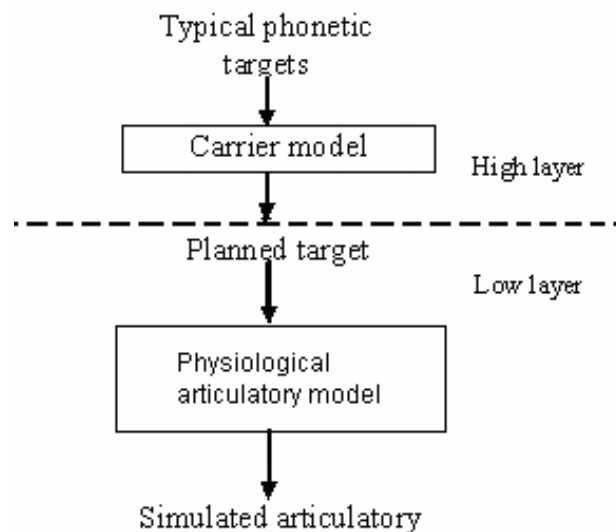


Figure 1.7 La production de la parole de modèle (DANG, 2005)

Le "carrier model" est bâti sur l'hypothèse selon laquelle l'articulation peut être séparée en deux ensembles, un mouvement vocalique et un mouvement consonantique, et que les consonnes se superposent à une base vocalique (Öhman, 1966). Le mouvement articulaire est alors modélisé comme un mouvement continu d'une voyelle à une autre, perturbé par le mouvement vers la cible consonantique (Figure 1.8) (les indices i et j sont utilisés pour la consonne et la voyelle respectivement).

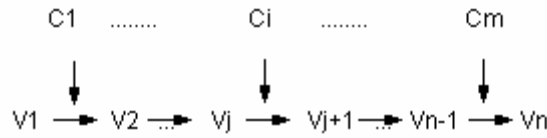


Fig1.8. Le principe de modélisation de cible et du mouvement pour les voyelles et pour les consonnes dans “carrier model”.

La base vocalique sur laquelle se superpose la commande consonantique est déterminée comme une pondération des cibles vocaliques. Chacune de ces cibles est elle-même pondérée par un coefficient intrinsèque et constant caractérisant la tolérance de la cible à la coarticulation, similaire au degré de contrainte articulatoire (DAC) proposé par Recasens (Recasens, 1997, 2002). Dang *et al.* (2005) ont proposé ainsi une formulation de la base vocalique sur laquelle se superpose la consonne C_i sous la forme

$$G_i = \frac{\alpha d_{V_j} V_j + \beta d_{V_{j+1}} V_{j+1}}{\alpha d_{V_j} + \beta d_{V_{j+1}}}$$

où V_j et V_{j+1} sont les voyelles adjacentes de la consonne C_i . d_{V_j} et $d_{V_{j+1}}$ sont les DAC des voyelles (plus le DAC est fort, moins la voyelle sera sensible aux effets de coarticulation). α et β sont les coefficients de pondération.

La position consonantique planifiée intègre cette base vocalique et la cible phonétique type de la consonne. Pour la consonne C_i , elle s'exprime alors :

$$C_i' = (r_{C_i} C_i + G_i) / (r_{C_i} + 1)$$

r_{C_i} traduisant la résistance à la coarticulation de C_i . Par rapport au modèle d'Öhman, Dang *et al.* introduisent une dimension supplémentaire. En effet, ils ont proposé que, tout comme les cibles phonétiques types des voyelles influencent la cible planifiée de la consonne, la cible planifiée de la consonne influence la cible planifiée pour la voyelle précédant la consonne. Ainsi pour la voyelle V_j , la cible planifiée devient

$$V_j' = (d_{C_i} C_i' + d_{V_j} V_j) / (d_{C_i} + d_{V_j})$$

d_{C_i} et d_{V_j} étant les DAC respectif de C_i et V_j .

Dang *et al.* proposent donc un modèle de planification qui concerne l'ensemble de la séquence VCV, dépassant donc le strict cadre syllabique. Les voyelles influencent les objectifs consonantiques qui, à leur tour, agissent sur les objectifs vocaliques. Il est aussi intéressant de constater que bien que la base vocalique soit le support de la planification de la cible

consonantique, la voyelle V_{j+1} n'influence pas directement la voyelle V_j , mais par le biais de la consonne intermédiaire. Il faut aussi noter que la coarticulation est seulement régressive pour la voyelle (seul V_j est influencée par les phonèmes voisins dans la séquence $V_j C_j V_{j+1}$) alors qu'elle est régressive et progressive pour la consonne influencée à la fois par V_j et V_{j+1} .

Dang *et al.* ne proposent pas de cadre théorique pour déterminer α et β , d_{cj} , d_{vj} et r_{ci} , mais un cadre heuristique : ils proposent d'optimiser ces coefficients de façon à décrire au mieux les données observées (des données EMMA du japonais). Il s'agit donc d'une formalisation mathématique offrant un cadre efficace pour analyser et synthétiser des données expérimentales avec un modèle physiologique du conduit vocal pour la production de parole.

1.3.2 "Task-dynamics" et "Coordinative structure" : Le modèle des laboratoires Haskins

Dans le modèle élaboré par les laboratoires Haskins, les unités phonologiques sont de nature gestuelle conformément aux propositions de la Phonologie Articulatoire élaborée par Browman et Goldstein (1989). La coarticulation y résulte alors de la coproduction des gestes spécifiés par la structure phonologique sous-jacente (la constellation de geste) et ceci sous l'influence de 2 caractéristiques des gestes : (1) leurs propriétés dynamiques et celles de leurs interactions ("Task-dynamics") (Saltzman et Munhall, 1989 ; Kelso *et al.*, 1986). (2) et les coordinations entre les articulateurs recrutés pour chaque geste (les "Coordinative structure") (Turvey, 1977 ; Saltzman, 1986 ; Kelso *et al.*, 1984 ; Fowler, 1980).

Les objectifs des gestes sont spécifiés dans l'espace des variables du conduit vocal ("*vocal tract variable*") sous forme d'attracteurs dynamiques. (en l'occurrence des systèmes masse ressort). Les variables du conduit vocal sont les différents lieux de constriction du conduit vocal (dans la zone du corps de la langue : Tongue body " T_B ", dans la zone apicale "Apex" et dans la zone labiale "Lips", l'ouverture du post velopharyngé "VEL" et l'ouverture glottique "GLO"). (Figure 1.9). La production d'un son est alors associée à l'activation plus ou moins forte d'un ou plusieurs de ces attracteurs (Kelso *et al.*, 1986) (Figure, 1.10). L'activation coordonnée de ces différents attracteurs génère la coproduction coordonnée de gestes, ce qui va définir des trajectoires dans l'espace des tâches dont les décours temporels dépendent du caractère dynamique de chacun de ces gestes et de leurs interactions (cf. Figure 1.9 et 1.11). La façon dont les articulateurs du conduit vocal (la langue, la mandibule, les lèvres) se déplacent au cours du temps va dépendre du mécanisme de coordination entre ces articulateurs tels qu'ils sont décrits pour chaque geste dans les structures coordinatrices.

Le niveau de coordination inter-gestuel spécifie la synchronisation et la cohésion parmi

des patrons d'activation des unités gestuelles qui participent à la production d'une séquence donnée. Par exemple, l'intervalle d'activation pour le geste du corps de la langue et le geste bilabial dans une séquence voyelle-labial-voyelle. Le niveau de coordination entre articulateurs explique la coordination parmi des articulateurs à un moment indiqué à l'ensemble d'active de geste, telle que, par exemple la coordination entre les lèvres, la mandibule et la langue pendant la production d'une constriction du corps de la langue T_B et d'un geste bilabial.

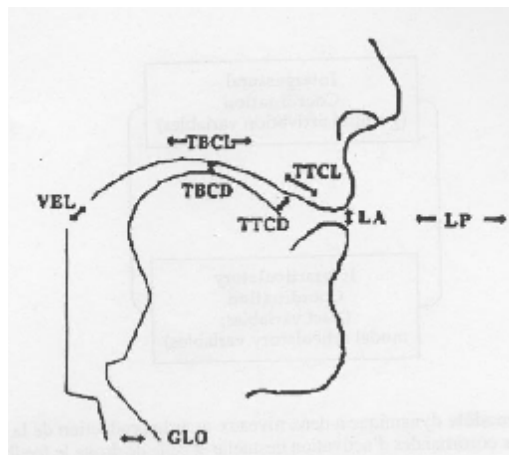


Figure 1.9 Les variables du conduit vocal (positions et degrés de constrictions). Saltzman et Munhall (1989)

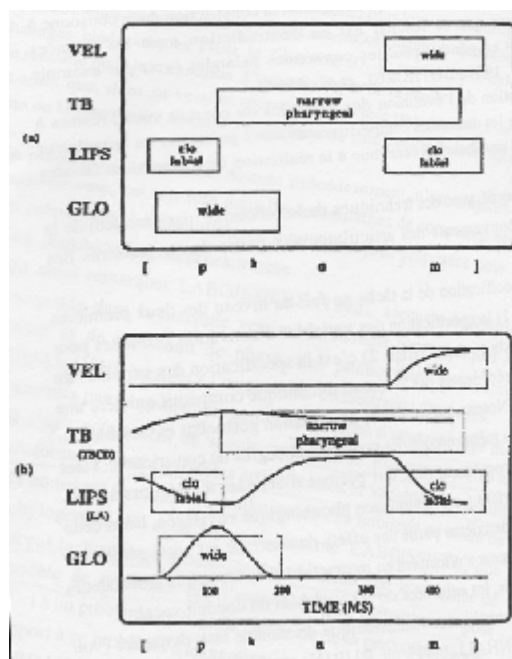


Figure 1.10 Réalisation du mot anglais « palm » (a) Intervalles d'activation des variables gestuelles, (b) Trajectoires du second ordre calculé. Saltzman et Munhall (1989)

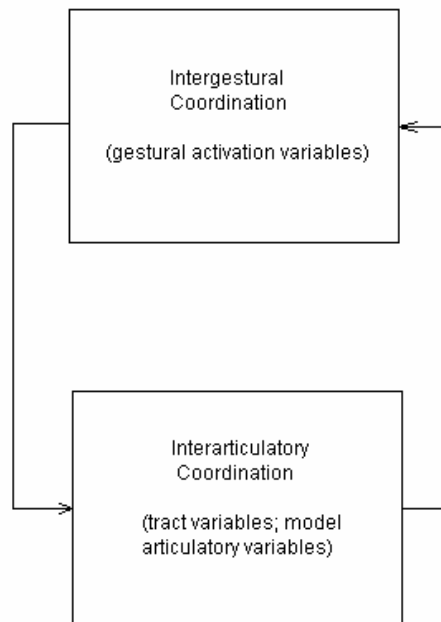


Figure 1.11 Le modèle la dynamique de la tâche (“*task dynamics*”) à deux niveaux pour la production de la parole. La flèche à gauche représente les commandes d’activation gestuelles et la flèche à droite représente le feedback de l’état du système physique (Saltzman et Munhall 1989)

Ainsi le déroulement temporel des articulateurs dépend des propriétés dynamiques des attracteurs caractérisant le geste, de la dynamique des interactions gestuelles, du timing de la participation gestuelle, et de la coordination entre articulateurs spécifiés par les structures coordinatrices.

1.3.3 Planification optimale

1.3.3.1 Degrés de liberté en excès et réseaux séquentiels

L’idée d’optimalité des stratégies de coarticulation est fondée sur l’exploitation de la non spécification de certains traits, telle qu’elle a été proposée par Keating dans son modèle de fenêtre (Keating, 1998). Cette idée a rejoint les propositions d’optimalité des stratégies motrices exploitant les degrés de liberté en excès du système moteur, élaborées à la même époque dans le domaine de l’étude et de la modélisation des gestes humains précis (cf par exemple Rumelhart et Norman, 1982). Ces propositions proviennent d’un double constat paradoxal : (1) dessiner, pointer avec le doigt, parler, sont des tâches motrices qui se réalisent dans un espace physique dont la dimensionnalité est plus faible que celle de l’espace de la commande. Le mouvement du doigt dans une tâche de pointage s’exprime ainsi dans l’espace géométrique tridimensionnel alors que sa position dépend de l’angle des articulations de l’épaule, du coude, du poignet, du doigt, des phalanges...; Ces degrés de liberté en excès

offrent de nombreuses possibilités de commande pour une même trajectoire du doigt. (2) pourtant, la répétition d'un même geste précis par une même personne dans les mêmes conditions montre une régularité très forte des stratégies motrices sous-jacentes. Pour expliquer ce paradoxe, il a été proposé que l'élaboration des stratégies motrices pourrait répondre à une contrainte d'optimalité qui réduirait le nombre de solutions possibles (Jordan, 1990). Selon cette conception, la planification d'une tâche motrice précise consiste donc à chercher les séquences de commandes motrices en minimisant le coût sur l'ensemble de la séquence.

Kawato *et al.* (1987) et Jordan et Rumelhart (1992) ont proposé que ce processus d'optimisation fasse appel à des représentations internes du système moteur, des modèles dans le Système Nerveux Central (on parle de modèle interne), qui permettent de prédire les conséquences des commandes motrices sur le déplacement de système moteur. Selon Jordan et Rumelhart (1992), ces modèles internes seraient basés sur la notion de Jacobien, permettant de prédire la sensibilité des variables de sortie à des variations locales de commandes motrices, et sur la notion de Jacobien Inverse, permettant de savoir comment faire varier les commandes motrices pour obtenir une variation donnée des variables de sortie.

Différents coûts ont aussi été proposés (par exemple, Nelson, 1983 pour une caractérisation des conséquences de la minimisation de certains coûts, vitesse maximale, impulsion, jerk..., sur les caractéristiques cinématiques des signaux du mouvement). Kawato (1990) et collègues ont ainsi suggéré que la force totale ou le couple total soit minimisés sur l'ensemble de la séquence.

Jordan (1989) a proposé des stratégies qui consistent à rechercher des solutions garantissant l'évolution la plus lisse des paramètres de sortie ("maximal smoothness"). Jordan a travaillé sur un certain nombre d'applications concrètes de ces hypothèses pour la production de la parole. Il a aussi par exemple cherché à trouver des stratégies d'anticipation observées par Benguerel et Cowan (1974). Il a utilisé pour cela un réseau de neurones séquentiel dont les variables de sortie correspondent à 8 traits articulatoires. Cette caractérisation va dans le sens des propositions de Henke (1966) ou Keating (1988b). L'entrée du réseau est un "plan d'action" qui spécifie les valeurs de ces 8 traits à des instants spécifiques de la séquence. Conformément à ce qu'avaient suggéré Henke (1966) et Keating (1988b), tous les traits n'ont pas à être spécifiés à chacun de ces instants, ce qui offre une certaine liberté grâce à laquelle une stratégie d'optimisation peut être mise en place. Elle va consister à définir pour chacun des instants spécifiés dans le plan d'action l'intégralité des

traits en recherchant la variation la plus lisse de chacun d'entre eux. Ces valeurs cibles sont alors considérées comme des attracteurs, au sens de la dynamique de la tâche de Saltzman (1989), qui va déterminer l'évolution temporelle des traits entre chacune d'elles. Les observations de Benguerel et Cowan (1974) ont ainsi effectivement pu être bien reproduites.

Ce premier exemple ne nécessitait pas de représentations internes puisque les paramètres de sortie correspondaient directement aux paramètres d'entrée. Une exploitation de ces mêmes principes a été faite par la suite pour le contrôle d'un modèle articulatoire du conduit vocal où les sorties étaient des trajectoires formantiques et les entrées des paramètres de commande du modèle articulatoire. Cela impliquait cette fois-ci l'utilisation de modèles internes (Jacobien et Jacobien inverse) des relations articulatoire-formantiques (Bailly *et al.*, 1990). Ici, des phénomènes de coarticulation conformes à des données recueillies sur des locuteurs, ont pu de nouveau être mis en évidence, en recherchant les solutions correspondant aux variations les plus lisses des commandes articulatoires.

Cette notion de planification optimale a été aussi utilisée par Vatikiotis-Bateson *et al.* (1994), dans un cadre similaire proposé par Kawato (1990). Notons cependant que pour le contrôle moteur en général, les hypothèses de modélisation de Kawato font appel à des modèles internes plus complexes capables de résoudre des problèmes d'inversion dynamiques dans le but d'optimiser des critères dynamiques (force, couple) et non plus seulement cinématiques (comme le lissage maximal).

1.3.3.2 Le modèle DIVA : l'optimalité dans un cadre syllabique

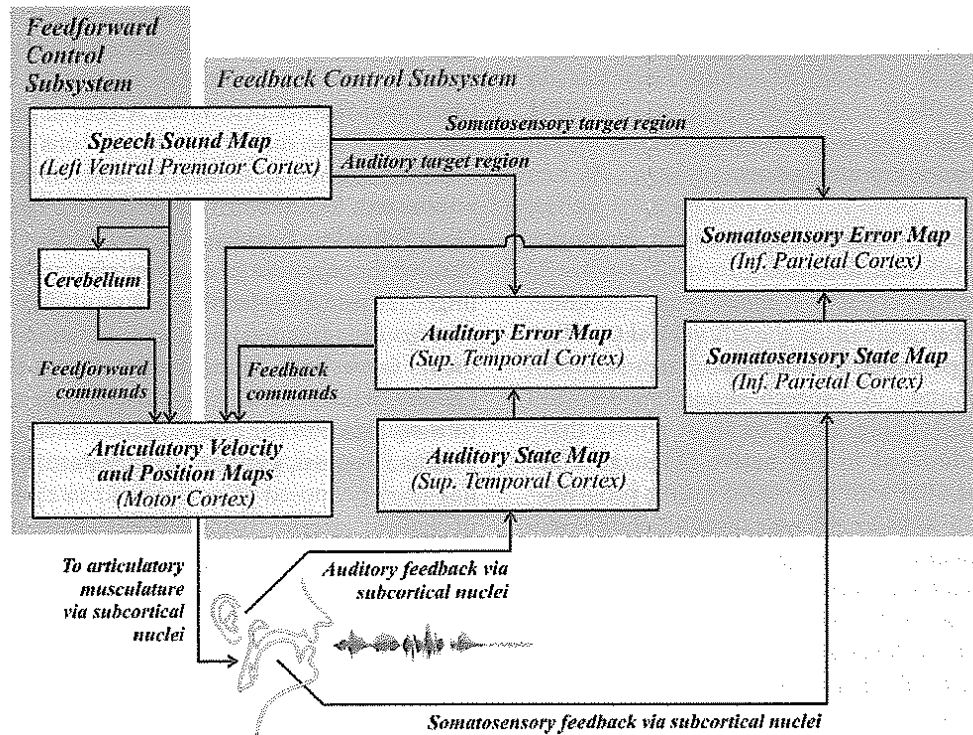


Fig 1.12. Modèle DIVA de production de la parole (Guenther 2006)

Un des modèles de production de la parole les plus en vogue est le modèle DIVA proposé par Guenther (Guenther, 1995 ; Guenther *et al.*, 1998 ; Guenther *et al.*, 2006) (Figure. 1.12). Dans ce modèle, les sons sont définis dans un espace bimodal associant caractéristiques auditives (formants) et informations somato-sensorielles (production de la constriction dans le conduit vocal). Ils sont caractérisés soit par des régions cibles (phonèmes statiques) soit par des régions spatiotemporelles cibles dans ces espaces (diphthongues ou syllabes). Ces informations caractéristiques sont stockées dans une première courte sensorielle (“Speech Sound Map”). L’apprentissage de la parole permet d’associer à la fois des objectifs cibles à des régions de l’espace articulaire, et pour chaque son, des informations sur les déplacements articulaires nécessaires pour atteindre les caractéristiques auditives et somato-sensorielles d’un autre son. Ces commandes articulaires sont apprises au cours du babillage. Elles sont stockées dans une autre carte sensorielle (“Articulatory Velocity and Position Maps”).

Pour chaque son, les changements articulaires nécessaires pour passer à un autre son sont appris grâce à la recherche d’une solution optimale qui minimise un double critère : la distance entre les caractéristiques auditives et orosensorielles produits et celles qui sont

spécifiées pour le son à atteindre, et l'écart entre les commandes articulatoires obtenues et celles de la configuration située au centre de la région cible. Ainsi, on obtient la meilleure réalisation possible des objectifs auditifs et orosensoriels tout en ayant les configurations articulatoires les plus "canoniques". Pour les diphtongues ou pour les syllabes les plus fréquentes, l'évolution des commandes articulatoires, permettant une meilleure reproduction des cibles spatiotemporelles, est également stockée. Cette hypothèse est compatible avec l'idée d'un syllabaire mental ("*mental syllabary*") proposé par Levelt (1994) selon laquelle un locuteur adulte aurait stocké, au niveau cortical, les patrons articulatoires associés à la production de syllabes.

L'exécution des mouvements articulatoires correspondant à une syllabe se fait alors en appliquant les changements articulatoires spécifiés par la deuxième carte sensorielle aux articulateurs du conduit vocal. Ceci se fait sous le contrôle d'un contrôleur feedback qui peut venir modifier les changements articulatoires recrutés en fonction de l'évaluation de l'écart entre les caractéristiques auditives et somato-sensorielles attendues et les caractéristiques produites, via des cartes sensorielles inverses ("Auditory Error Map") et ("Somatosensory Error Map"). Les caractéristiques produites ne sont en réalité pas celles des signaux articulatoires et acoustiques effectivement générées par le système physique de production de la parole, mais les sorties de modèle interne simulant les relations entre changements articulatoires et caractéristiques auditives ou somato-sensorielles ("Auditory state Map") et ("Somatosensory State Map"). Pour une séquence de parole supérieure à la syllabe, le processus d'exécution consiste à juxtaposer les séquences articulatoires associées à chaque syllabe.

On voit donc que la planification motrice optimale de DIVA semble à l'image des propositions de Kozhevnikov, V. et Chistovich, L. (1965) se limiter à la syllabe qui est considérée comme l'unité phonologique structurante prépondérante de la parole.

Chapitre 2 : Description phonologique simplifiée et coarticulation pour le chinois et pour le français

Introduction

La coarticulation n'est pas un phénomène universel. Elle est probablement influencée par la phonologie spécifique à chaque langue. Zerling (1991) a indiqué que, au niveau supérieur de la programmation motrice d'une chaîne de sons, les contraintes articulatoires pourraient surtout être imposées par les règles phonologiques de la langue.

Les différences dans la structure phonologique des langues ont été proposées pour expliquer les stratégies de production de la parole différentes selon les langues. Lubker et Gay (1982) ont étudié l'arrondissement des lèvres en suédois et en anglais américain. Ils ont caractérisé la variabilité des stratégies propres à chaque langue. L'anticipation labiale semble ainsi être spécifique à chaque langue. Ces auteurs ont observé que le suédois est plus exigeant sur l'arrondissement des lèvres que l'anglais américain. Cela peut être expliqué par le système vocalique du suédois qui est plus riche que le système vocalique anglais. Le geste de protrusion des lèvres serait alors plus net pour les locuteurs suédois parce qu'il leur faut assurer une différence suffisamment grande entre voyelles arrondies et non arrondies.

Hoole et Nguyen (1993) ont comparé l'effet de la coarticulation pour trois langues : l'allemand, le français et l'anglais. Les données articulatoires EPG ont été enregistrées par deux locuteurs pour chacune de ces langues. L'analyse de la coarticulation met l'accent sur les consonnes fricatives. Cette étude a contribué à soutenir le concept de dominance dans la direction de propagation de la coarticulation dans différentes langues. Les résultats montrent que les locuteurs allemands et anglais ont un effet prédominant pour le carryover par rapport à l'anticipation. En français, l'un des deux locuteurs préfère l'effet d'anticipation sur l'effet du carryover. L'autre ne montre aucune préférence entre l'effet d'anticipation et de carryover.

Dans le projet ACCOR dirigé par Marchal et Hardcastle (1993), le processus articulatoire spécifique à sept langues européennes a été examiné avec pour but de déterminer comment ce processus se différencie en fonction des différents systèmes phonologiques. Les sept langues ont leur propre système de voyelles et de consonnes. La structure de chaque système spécifique pourrait contraindre le type et l'étendue de la coarticulation dans la langue concernée. Marchal (1988) a indiqué l'importance de travailler sur la coarticulation à travers différentes langues : *“It's obviously true that coarticulation does not always apply to different languages in the same way (Perkell, 1986). Coarticulation should thus be investigated in*

different languages. We could then experimentally determine what effects are caused by purely mechanical constraints and are therefore universal and what effects are language-specific, expressing the phonological conditions of the speech production process.”(p293)

La recherche de la coarticulation à travers plusieurs langues (cross-language) peut distinguer les caractéristiques spécifiques à chaque langue et exploiter les différentes stratégies employées par locuteur.

Le but de notre travail est d'étudier les effets potentiels de la structure phonologique sur les stratégies de planification dans la coarticulation en chinois et en français. Le choix du français et du chinois est judicieux pour les objectifs de notre étude, car ils possèdent des systèmes phonologiques différents, en particulier du point de vue du statut de la syllabe. Dans ce chapitre, nous présenterons dans un premier temps une description phonologique pour le chinois et pour le français. Dans cette description phonologique, nous nous intéressons notamment à l'inventaire vocalique et au statut syllabique qui peuvent être relatifs à la coarticulation de chacune des deux langues. Cependant nous devons rappeler que le chinois dont nous parlons dans cette étude est le chinois mandarin. Le mandarin est membre de la famille des langues nordiques en chine. Il est basé sur la prononciation du dialecte de Pékin. Le mandarin est la langue officielle et la langue la plus parlée en Chine. Après la description phonologique simplifiée respective pour le chinois et pour le français, nous présentons quelques travaux effectués sur la coarticulation à partir de données acoustiques et articulatoires dans ces deux langues.

2.1 Description phonologique simplifiée pour le chinois et pour le français

2.1.1 L'inventaire vocalique

Il a été suggéré que la densité de l'inventaire vocalique d'une langue influence la variabilité de voyelle liée à la coarticulation de voyelle à voyelle (Manuel, 1984, 1990). Manuel a trouvé qu'une langue avec un espace vocalique peu dense permet plus d'effets de coarticulation de voyelle à voyelle. En revanche, une langue qui a un espace plus dense montre moins d'effet de coarticulation. Manuel a pour cela étudié la coarticulation voyelle à voyelle sur des données acoustiques pour différentes langues qui ont un nombre différent de voyelle dans leur système phonologique (Manuel et Krakow, 1984). Les résultats montrent que l'effet de coarticulation de voyelle à voyelle est plus grand et plus étendu pour le swahili et le shona (système à cinq voyelles) que pour l'anglais (système à 13-15 voyelles orales).

2.1.1.1 L'inventaire vocalique en chinois

La description du système des voyelles du chinois a été et est toujours très controversée. Selon une description phonologique classique, le chinois mandarin a principalement cinq voyelles, /i, y, u, ə, a/ (Chao, 1968 ; Deng et Dang, 2007). Cependant, la situation est plus complexe dans les vrais environnements phonétiques, où un certain nombre d'allophones émergent dans différents contextes phonétiques. La voyelle antérieure non arrondie /i/ varie entre /i/, /ɨ/ et /ɪ/ selon le contexte consonantique. La voyelle /ɨ/ est réalisée quand la consonne précédente est un /s/ un /ts/ ou un /ts^h/. La voyelle /ɪ/ est employée quand la consonne précédente est un /ʃ/ un /tʃ/ ou un /tʃ^h/. La voyelle centrale /ə/ peut être une mi centrale /ə/, une mi antérieur /e/, une mi postérieur /ɤ/ ou une arrondie /o/ (Chao, 1968 ; Duanmu, 2002). La voyelle basse /a/ peut se réaliser comme /a/ avant une voyelle /u/ et comme /æ/ avant une consonne /n/ (Duanmu, 2002). Mais phonologiquement il s'agit de variantes allophoniques et non de vraies classes phonologiques. Par conséquent, phonologiquement le chinois mandarin est considéré comme ayant seulement cinq voyelles, mais phonétiquement il a dix réalisations de voyelles. Ces réalisations sont associées avec quatre endroits d'articulation (antérieur, mi-antérieur, mi-postérieur et postérieur), trois degrés d'ouverture (haute, mi, basse), et deux caractéristiques d'arrondissement pour les voyelles antérieures (arrondies et non arrondies).

2.1.1.2 L'inventaire vocalique en français

L'inventaire vocalique français est décrit comme ayant 11 voyelles monophthongues /i, u, y, e, o, ø, ɛ, œ, ɔ, ɑ, a/ et quatre voyelles nasales /ẽ, ã, œ̃, õ/ (Calliope, 1989). La distinction parmi les 11 voyelles orales est réalisée dans un espace tridimensionnel, en prenant trois lieux d'articulation (de l'avant à l'arrière), quatre degrés d'ouverture dans le conduit vocal et les caractéristiques d'arrondissement des lèvres pour des voyelles antérieures (arrondies et non arrondies). L'opposition entre la voyelle antérieure [a] et la voyelle postérieure [ɑ] est limitée à certains accents du Nord. Les paires [e]/[ɛ], [ø]/[œ], [o]/[ɔ] sont liées à des caractéristiques d'ouverture. En d'autres termes, elles se distinguent par une réalisation fermée ou ouverte.

2.1.2 Le rôle de la syllabe

Selon l'hypothèse de Fowler (1980), le niveau supérieur du contrôle de la parole est responsable de tout ce qui est structuration des entités abstraites phonologiques. Les problèmes essentiels de cette structuration concernent le choix de l'unité de base dans lequel les entités abstraites seraient encadrées. Le choix de l'unité de base soulève le problème de la représentation et de la mémorisation dans le cerveau. La syllabe est devenue un choix pour l'unité de base linguistique. Levelt (1994) a proposé que pendant la production de la parole, le locuteur stocke mentalement les syllabes dans le cerveau. Le syllabaire mental a été défini comme un mécanisme pour traduire une représentation phonologique abstraite en un plan phonétique en fonction du contexte.

“The mental syllabary was postulated as a mechanism for translating an abstract phonological representation of an utterance into a context-dependent phonetic representation which is detailed enough to guide articulation.” Levelt (1994)

Le syllabaire mental est un programme de stockage syllabique au niveau de la représentation phonétique. Certaines théories ont proposé une pré-articulation qui forme seulement une représentation au niveau phonologique en composant des segments directs, et suggéré que ce niveau de représentation active directement l'articulation (Mackay, 1982). De plus Levelt a proposé que, à part la représentation phonologique, une représentation phonétique dépendant du contexte est nécessaire, car la réalisation phonétique d'un segment phonologique est déterminée par le contexte dans la parole. Ces effets phonétiques pourraient être différents selon la langue. Ils ne peuvent pas tous être produits par l'implémentation des règles phonétiques universelles. De tels effets peuvent être seulement le résultat de la représentation phonétique propre à une langue. Dans l'hypothèse de syllabaire mental de Levelt, les entités abstraites phonologiques forment chaque unité syllabique dans l'étape de représentation phonétique. Cette représentation phonétique est employée pour guider directement le système articulatoire. La syllabe est commandée comme une unité entière pour l'articulation.

La syllabe semble aussi jouer un rôle crucial dans la segmentation du flux de la parole. Cette segmentation recouvre des processus employés par l'auditeur pour convertir le signal de la parole en une séquence de mots. Frauenfelder et Nguyen (2000) ont proposé une représentation mentale (“infra-lexicale”) qui projette le signal de la parole en une séquence

phonétique. Le problème de la segmentation dans la parole est lié au fait que l'unité de segmentation peut être le phonème, la syllabe ou le mot. Mehler (1981,1990) a proposé que la parole soit segmentée et classifiée par les syllabes.

Selon ces auteurs, un mécanisme de haute niveau a été proposé pour la production et la perception de la parole. Cependant le rôle de la syllabe semble être différent selon les langues. Cette représentation d'une langue pourrait guider l'articulation de manière différente. C'est pourquoi nous nous sommes respectivement intéressées au statut de la syllabe pour le chinois et pour le français.

2.1.2.1 La syllabe en chinois

En chinois, la syllabe est classiquement considérée comme l'unité structurante phonologique la plus importante. C'est une caractéristique très forte de cette langue. En effet, en tant que langue écrite, elle s'appuie sur des caractères (idéogrammes) correspondant systématiquement à une syllabe. Cette description a ainsi fortement influencé des travaux sur la reconnaissance de la parole, la synthèse de la parole, et la recherche documentaire pour le chinois. De plus, chaque morphème chinois correspond généralement à une syllabe. Une syllabe peut également fonctionner en tant que mot monosyllabique. Certains auteurs pensent que le chinois est une langue monosyllabique (Karlgren, 1949), mais d'autres pensent que le chinois est dissyllabique (Duanmu, 2000). En fait, un mot chinois peut être monosyllabique, dissyllabique ou trisyllabique. Mais une étude plus approfondie du lexique chinois révèle que le mot dissyllabique prend le dessus dans le vocabulaire moderne (enquête menée par l'organisation *Chinese Language Reform Committee Research and Popularization Office*, 1959). Dans le mot dissyllabique, deux syllabes qui correspondent à deux caractères chinois fonctionnent en qualité de deux morphèmes.

La raison principale pour laquelle le chinois est considéré comme une langue à base syllabique provient du fait qu'il est une langue à tons. Chaque ton est superposé à une syllabe. Le chinois attribue ainsi à la syllabe un rôle majeur puisque c'est elle qui décide du positionnement et de la durée des tons. Il y a quatre tons distinctifs, portant toujours sur la syllabe forte, ce sont : (1) le ton haut ; (2) le ton montant ; (3) le ton bas ; (4) le ton descendant. La syllabe forte correspond essentiellement au mot monosyllabique de contenu (monosyllabic content words). Parmi toutes les syllabes chinoises, la plupart sont des syllabes fortes, pour lesquelles les quatre tons sont assez uniformément distribués. La plupart des syllabes ont quatre ou trois tons chacune, et seulement un petit nombre de syllabes en a

deux ou un. C'est principalement, la syllabe faible qui n'a pas son propre ton, mais elle peut obtenir le ton dans un certain environnement d'intonation. Les tons concernent toujours une syllabe dans sa totalité. Ils ne traversent pas les frontières de la syllabe. Nous présentons un exemple de caractères, de syllabes, et de tons en chinois : Quatre caractères chinois 妈, 麻, 马, 骂 correspondent à une seule syllabe /ma/. Ils se distinguent l'un de l'autre par quatre tons différents /ma¹/, /ma²/, /ma³/, /ma⁴/.

Une autre raison qui montre l'importance du statut de la syllabe en chinois est que, dans cette langue, les syllabes ne sont pas simplement la combinaison de phonèmes simples. En effet, de nombreuses règles déterminent la structure des syllabes du chinois. Il y a 22 consonnes. Contrairement à la langue occidentale, la distribution des consonnes en chinois dépend plus étroitement de la position de syllabe dans le mot, et la structure de syllabe est beaucoup plus simple (Deng et Dang, 2007). Notamment, il n'y a pas de cluster de consonnes en chinois. Seule la consonne simple peut apparaître dans une syllabe à l'attaque ou en coda. Cependant il existe des contraintes supplémentaires : /ŋ/ ne peut pas figurer à l'attaque et seules les nasales /n/ ou /ŋ/ peuvent apparaître en position coda. Cela revient à dire qu'aucune syllabe ne peut commencer par une consonne /ŋ/ et qu'aucune syllabe ne peut finir par une consonne /t/ ou une /m/ ou une /l/ par exemple.

Une conséquence de ces contraintes fortes peut s'expliquer par le fait que, contrairement à l'anglais ou au français, qui comptent des milliers de syllabes, le chinois ne possède qu'environ 400 syllabes à l'exclusion des tons et 1300 si on inclut les tons (Deng et Dang, 2007). Selon un corpus de textes de plus de 45 millions de caractères chinois (une enquête de Da sur "*Chinese text computing*", 2000), il existe plus de 6000 caractères différents. Cela signifie que chaque syllabe chinoise représente environ 15 caractères à l'exclusion des tons, ou 5 caractères si les tons sont inclus. La distribution de ces homophones n'est pas uniforme. La densité d'homophones en chinois analysés pour 6000 caractères chinois par Da (2000) montre que la plupart des syllabes représentent moins de 20 caractères chacune, mais que certaines syllabes représentent environ 100 caractères (Duanmu, 2006). Par exemple, la syllabe /tɕi/ correspond à tous ces caractères chinois: 及, 即, 既, 几, 机, 济, 基, 记, 计, 技, 际, 己, 集, 级, 极, 纪, 积, 继, 击, 激, 急, 绩, 季, 吉, 疾, 迹, 给, 鸡, 剂, 籍, 辑, 寄, 肌, 挤, 圾, 忌, 稽, 其, 寂, 脊, 缉, 祭, 饥, 冀, 妓, 畸, 姬, 棘, 悸, 汲, 嫉, 系, 讷, 伎, 藉, 箕, 蓟, 奇, 期, 暨, 矾, 跻,

亟, 鲫, 羈, 稷, 驥, 瘠, 叽, 戟, 岌, 玑, 嵇, 劓, 髻, 芥, 楫, 霁, 唧, 畿, 笈, 芟, 犄, 佶, 嶠, 乩, 掎, 诘, 係, 覬, 蒺, 屣, 麇, 咭, 戢, 蹠, 虬, 鯨, 哿... Ils se distinguent par différents tons mais bien sûr surtout par le contexte.

On se rend compte ainsi de la force de la syllabe en chinois : (1) Elle n'est pas le résultat d'une combinaison généralisée entre tous les phonèmes et tous les disponibles. (2) Elle n'a pas besoin de se combiner avec d'autres pour être porteuse de sens linguistiquement parlant. La syllabe en chinois est donc très répétitive et peu variée. Ainsi, la syllabe en chinois semblent obéir à des règles très spécifiques et exister dans la structure linguistique de la langue plus comme une structure entière que comme une combinaison de phonèmes simples.

La phonologie traditionnelle du chinois considère deux unités distinctes au sein de la syllabe (Wu, 2000): une unité initiale correspondant à une consonne ou rien, et une unité finale correspondant à une voyelle ou une séquence de voyelles parfois accompagnée (s) d'une finale nasale. Il existe donc 5 types de position finale en chinois : une monophthongue V, une diphtongue VV, une triphthongue VVV, une monophthongue V+consonne nasale N, ou une diphtongue VV +consonne nasale N (N représente /n/ ou /ŋ/).

La colonne de gauche du tableau 2.1 montre la structure syllabique en chinois telle que la décrit la phonologie occidentale, c'est à dire avec une attaque, un noyau et un coda. Nous pouvons voir dans ce tableau que le chinois n'accepte pas de cluster consonantique, et que le coda peut seulement être une /n/ ou une /ŋ/.

Syllabe chinoise			Syllabe française		
Attaque	Rime		Attaque	Rime	
	Noyau	Coda		Noyau	Coda
Ø ou C	V (y compris diphtongues et triphthongues)	[n] ou [ŋ] Ø	Ø ou (CC)C	V	Ø ou C(CCC)

Tableau 2.1 : La structure syllabique en chinois à gauche et en français à droite
/Ø/ rien ; /C/ tout type de consonnes ; /V/ tout type de voyelles

2.1.2.2 La syllabe en français

a. “Resyllabification”

Le français est également classiquement considéré comme une langue à base syllabique. Un certain nombre de phénomènes a permis d'étayer cette hypothèse. Notamment, il existe en français un phénomène de “resyllabification” capable de transgresser les frontières de mots, donnant à la syllabe un statut plus fort que celui des unités lexicales. C'est ce qui a montré Delattre pour des séquences VCV en français, où toutes les consonnes intervocaliques se rattachent à la voyelle qui les suit, même si cette voyelle appartient à un autre mot. Ainsi, pour les séquences « too late » et « tool eight » en anglais, un français proposerait une seule forme de syllabification « too late » [tu] [let] alors qu'un anglais en proposerait deux : « tool late » [tu] [let] et « tool eight » [tul] [et] (Delattre, 1953). Delattre a rapporté également que les phrases « *elle imite un autre accent* » et « *elle limite un autre accent* » sont segmentées de la même façon : [ɛ-li-mi-tœ-no-tra-ksã].

Delattre (1940), a ainsi suggéré qu'en français la segmentation en mots serait difficile, et pourrait être masquée par la cohésion des syllabes aux frontières des mots. Il a indiqué que “ *L'élision, la liaison et l'enchaînement aidant, le groupe rythmique lui donne l'impression d'une succession de syllabes sensiblement égales dont le rythme ininterrompu cache malicieusement les limites des mots.* ”

M. Grammont (1938) a mentionné cet aspect du français, d'un point de vue pratique :

“*Quelqu'un qui ne sait pas où commencent et où finissent les mots français ne pourrait jamais le deviner en entendant parler. Quand on s'arrête, c'est après un mot, parce qu'avec ce mot l'idée, où une parcelle de l'idée, est terminée, mais le plus souvent aucun fait matériel ne marque la fin ou le commencement des mots. D'ordinaire les mots se disent par groupes, par séries, sans aucun arrêt, et si étroitement unis l'un à l'autre qu'il n'est pas rare qu'une syllabe soit constituée par la fin d'un mot et le commencement d'un autre.* ”

Du fait de la “resyllabification”, il existe de nombreuses syllabes ouvertes en français. Delattre (1969) a fait une étude phonologique sur la structure de la syllabe en anglais, en allemand, en français et en espagnol. Le caractère syllabique est souligné par la comparaison de 4 langues. 2000 syllabes y ont été analysées dans chacune des quatre langues. Il est indiqué que les syllabes fermées sont prédominantes dans la langue allemande et anglaise (63% et 60%). Pour le français et l'espagnol, les syllabes ouvertes sont plus fréquentes (76% pour le

français et 72% pour l'espagnol). Dans tous les corpus, les syllabes en français et en espagnol sont divisées indépendamment des frontières de mot dans le groupe de sens, mais en allemand et en anglais les consonnes finales de mot sont maintenues avec les dernières syllabes du mot. En ce qui concerne la variété de la structure syllabique, Delattre a trouvé 4 structures de haute fréquence pour les quatre langues: CVC, CV, VC et CCV. Les structures prédominantes sont CVC et CV en anglais et en allemand, tandis qu'en français et en espagnol la structure prédominante est CV. Le nombre de syllabes différentes pour chaque structure syllabique (une syllabe étant représentée ici phonétiquement) est aussi très variable en fonction de la langue. Delattre a rapporté que l'espagnol montre peu de variations dans les différentes syllabes utilisées, alors qu'une variété beaucoup plus grande est observée en anglais, en français et en allemand. En résumé de cette étude, pour le français, la structure syllabique prédominante est CV et les syllabes sont très variées phonétiquement.

b. Le rôle de la syllabe pour la segmentation en français

En reprenant l'hypothèse de Mehler sur le rôle important de la syllabe pour la segmentation en français, nous allons présenter quelques travaux sur le statut de la syllabe en français. Mehler (1981, 1990) a proposé que la syllabe constitue l'unité de base pour la classification dans la perception de la parole et dans l'accès au lexique. Particulièrement, il a indiqué que le signal de la parole pourrait être recodé et catégorisé prélexiquement sous forme d'unités syllabiques.

“Speech is segmented into elementary units that roughly correspond to the syllable. . . . Syllabic frames are recognized by a bank of syllabic analyzers . . .” (Mehler, 1990, p. 255).

Les auditeurs français sont plus rapides à la détection de séquences pour les cibles (une cible est un protocole expérimental pour reconnaître une syllabe) qui correspondent bien à la première syllabe (par exemple, /ba/ dans /balance/ ou /bal/ dans /balcon/) que pour les cibles qui correspondent moins à la première syllabe (par exemple, /bal/ dans /balance/ ou /ba/ dans /balcon/). Cette interaction entre le type de cible (CV contre CVC) et la structure du mot (mots de CV contre mots de CVC) est appelée l'effet syllabique.

Mais une étude récente de perception du français a tempéré cette conclusion sur l'effet syllabique. Content *et al.* (2001) ont ainsi indiqué que l'effet syllabique pour la segmentation pourrait être d'abord lié aux propriétés acoustiques-phonétiques spécifiques de quelques types de consonnes (Content, Meunier, Kearns, et Frauenfelder, 2001a). Dans les données de Mehler *et al.* (1981), tous les mots commencent par une /l/ ou une /r/. Content *et al.* ont utilisé

des pseudos mots variés de structure syllabique variée et avec différentes consonnes pour la détection de séquence CV et CVC. Ils n'ont pas retrouvé l'effet syllabique. Une interaction claire était cependant observée pour les consonnes liquides. Ainsi il est suggéré que l'effet syllabique pourrait être limité à l'utilisation des consonnes liquides. Cette étude a réduit la portée théorique de l'effet syllabique en français. Content *et al.* ont montré que les auditeurs pouvaient en réalité baser leur segmentation sur des indices acoustiques aux frontières syllabiques. Ce résultat ne montre pas d'évidence directe en faveur de l'hypothèse selon laquelle la syllabe sert de code d'accès au lexique mental en français. En outre, une autre étude de perception sur la syllabification conteste l'hypothèse selon laquelle les auditeurs français auraient des intuitions claires sur les frontières de syllabe. Content *et al.* (2001b) ont indiqué que les auditeurs français ne resyllabifiaient pas les mots simples de manière uniforme. (Content, Kearns, et Frauenfelder, 2001b). Les participants ont dû répéter la première ou deuxième partie des mêmes mots du stimulus CVCV. Une dissociation claire est observée. La grande majorité de réponses de la deuxième partie a inclus la consonne intervocalique (par exemple, *lon* pour le mot *ballon*), mais les réponses concernant la première-partie se montrent identiques pour CV et CVC (par exemple, *ba* ou *bal* pour le mot *ballon*). D'après Content, l'observation des réponses sur la deuxième-syllabe pour lesquelles la syllabification a commencé presque toujours par la consonne intervocalique suggère que les onsets des syllabes constituent des points fiables de segmentation dans le signal. Content a indiqué que la syllabe en français est principalement impliquée dans la segmentation plutôt qu'utilisée comme unité de classification. Ces résultats montrent que la force de la syllabe en français au niveau de représentation est moins forte qu'elle était supposée à l'origine. La syllabe en français ne semble pas jouer un rôle aussi important qu'en chinois.

c. Le rôle de la syllabe pour la prosodie du français

Un autre point favorise l'idée d'une différence du statut de la syllabe entre le chinois et le français, c'est que le système d'intonation des deux langues est totalement différent. Le chinois est une langue à tons, et on a vu que la syllabe joue un rôle crucial dans le support de ton. En revanche, le français est une langue à groupe accentuel. À la différence d'autres langues romanes comme l'italien ou l'espagnol, l'intonation en français est caractérisée par une séquence des mouvements mélodiques montants qui délimitent des frontières de syntagme. Le lieu de l'accent est fixe au niveau du mot, mais sa réalisation dépend de la position d'un mot dans un syntagme (Grammont 1938 ; Delattre 1939). La dernière syllabe du

syntagme porte un accent primaire, et est souvent réalisée avec un mouvement mélodique montant.

Traditionnellement, en raison de sa position en fin de syntagme, l'accent primaire a été considéré comme accent du syntagme ou accent avec une nature syntaxique (Rossi, 1979 ; Martin, 1982). Le domaine de l'accent primaire dépend de nombreux facteurs tels que le rythme de la parole, la longueur du syntagme, ou la contrainte syntaxique et sémantique (Pasdeloup, 1990 ; Fougeron et Jun, 1998).

L'accent primaire a été différemment nommé selon les auteurs : “mot rythmique” by Pasdeloup (1990), “rhythmic unit” par Di Cristo et Hirst (1993), “group rythmique” par Delais-roussarie (1995), et “accentual phrase” par Jun et Fougeron (1995, 2000).

En plus de l'accent primaire, un syntagme en français a également un accent secondaire ou initial. À la différence de l'accent primaire, les corrélations phonétiques de l'accent initial sont moins claires. L'accent initial se produit la plupart du temps sur la première ou la deuxième syllabe d'un mot.

Jun et Fougeron (1995, 2002) ont proposé un modèle d'intonation du français. (cf. Figure 2.1)

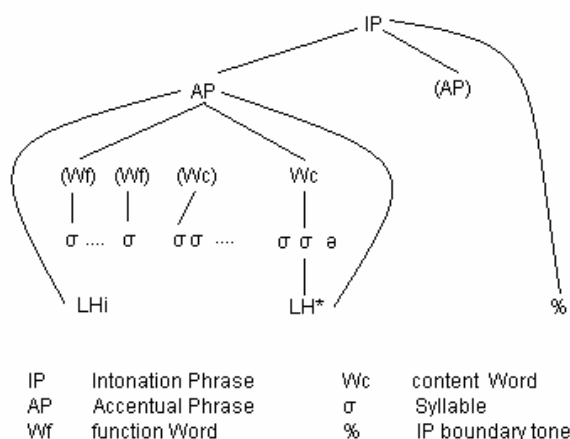


Figure 2.1 La structure de l'intonation en français. (Jun et Fougeron 2002)

AP est le domaine de l'accent primaire. Bien que le nombre de syllabes dans un AP varie soit selon la longueur du mot et la structure syntaxique ou sémantique du syntagme, soit selon des facteurs tels que le rythme de la parole et les locuteurs, Jun et Fougeron ont proposé qu'un AP contienne une moyenne de 2.3~2.6 mots, et 3.5~3.9 syllabes. Une unité de prosodie similaire a été proposée par Fonagy (1979) et par Pasdeloup (1990). Pour Fonagy (1979), l'unité “arc accentuel” contient également 3.36 syllabes en moyenne dans la conversation spontanée. Pour Pasdeloup (1990), l'unité “rythmique du mot ” contient en moyenne 3

syllabes avec un maximum de 7-8.

On voit donc qu'à la différence du chinois, l'unité prosodique est composée en français de plusieurs syllabes. Le chinois considère au contraire la syllabe individuelle comme une unité prosodique. Récemment, un modèle général d'organisation temporelle, a été proposé pour le chinois par Xu (2006). Dans ce modèle, la syllabe est considérée comme l'unité la plus fondamentale qui spécifie l'alignement F_0 des consonnes, des voyelles et des tons. La coarticulation se produit seulement entre la consonne initiale C et V dans la syllabe, il n'y a que peu ou pas d'anticipation de la consonne C sur V à travers la syllabe, aucune coarticulation de voyelle à voyelle à travers la consonne intervocalique. Comme cela a été dit plus haut, le statut de la syllabe, différent en français et en chinois, va vraisemblablement influencer les stratégies de coarticulation.

2.2 La coarticulation en français et en chinois

2.2.1 La coarticulation en français

L'articulation en français a été étudiée afin de connaître avec précision la forme du conduit vocal pour la production des voyelles et des consonnes, et pour analyser les mouvements réalisés au cours des transitions. La recherche de l'articulation s'est développée avec l'évolution des outils de mesure. Rousselot (1901) a créé un kymographe dès le début du 20ème siècle pour observer graphiquement l'articulation en français. Plus tard, les méthodes radiographiques ont été développées afin d'observer la position des articulateurs internes (la langue et larynx) pendant les tenues des voyelles et des consonnes (Straka, 1950 et Fouche, 1952). La cinéradiographie a été ensuite développée pour étudier les mouvements du conduit vocal, ainsi que les positions tenues (Brichler, 1970 ; Zerling, 1979). L'électropalatographie a été utilisée pour observer la palatalisation du français (Marchal, 1985; Nguyen et Marchal, 1993, 1996). Cette méthode a notamment permis de montrer les déplacements de la langue au cours de la phase d'occlusion. L'articulographie électromagnétique (EMMA) a été enfin développée pour mesurer les mouvements des articulateurs dans le plan mid-sagittal avec une résolution spatiale de l'ordre du millimètre. La vidéo a été aussi utilisée pour mesurer les mouvements labiaux dans l'étude de geste d'arrondissement (Aby et Lallouache, 1996).

2.2.1.1 La coarticulation d'arrondissement

La coarticulation d'arrondissement a été beaucoup discutée pour le français. Les principaux travaux dans ce domaine ont été déjà décrits dans le chapitre 1, comme certaines

applications majeures des théories princeps de la coarticulation. Nous en rappellerons donc seulement les grandes lignes. Le résultat connu de Benguerel et Cowan (1974) montre que la frontière syllabique ne constitue pas une limite à l'anticipation en français, contrairement à l'hypothèse formulée par Kozhevnikov et Chistovich (1965).

Abry et Lallouache (1996) ont étudié l'anticipation du mouvement de protrusion en français, notamment à travers les chaînes de consonnes relativement longues. Ils ont utilisé une phrase « *Ces deux Sixtes sculptèrent* », dans le cas où il existe deux voyelles /i...y/ autour du groupe de cinq consonnes /ktsk/. Les autres types de groupe de consonnes /ksk/, /kk/, /k/ sont aussi réalisés quand la chaîne de consonnes se réduit dans la phrase. En français, de tels groupes de cinq ou quatre consonnes se réalisent exclusivement à la jointure des mots. Les phrases utilisées dans cette étude sont donc constituées de groupes consonantiques à la jointure des mots. Les résultats montrent que le mouvement de protrusion de la voyelle /y/ commence de plus en plus tôt par rapport au début acoustique de la voyelle /y/ au fur et à mesure de l'augmentation du nombre de consonnes intervocaliques. Dans la chaîne de consonnes la plus longue /ktsk/, ce mouvement de protrusion commence après la voyelle non arrondie /i/. Dans la séquence d'une consonne /iky/, il peut commencer dès le début de la première voyelle non arrondie /i/. Cette étude montre que la frontière de la syllabe et la frontière du mot ne sont pas sensibles à l'anticipation d'arrondissement en français.

2.2.1.2 La coarticulation de nasalité en français

Benguerel (Benguerel et Hirose 1977a) a étudié la coarticulation vélaire en français par une étude fibroscopique. La trajectoire du vélum a été examinée sous l'angle de la coarticulation anticipatoire. L'anticipation de voyelles nasales sur les consonnes et l'anticipation de consonnes nasales sur les voyelles ont été étudiées. Pour l'anticipation de la voyelle dans les consonnes, la séquence est /rzfrV_n/ a été utilisée. La "V_n" représente une voyelle nasale précédée par un cluster consonantique non nasal. Benguerel a observé que le geste d'abaissement du vélum de la voyelle "V_n" commence pendant la deuxième consonne /z/. Vers la fin de la dernière consonne, le geste a atteint sa vitesse maximum vers le bas. Cette observation confirme les résultats obtenus par une autre étude similaire sur le flux d'air aux narines (Benguerel, 1974), dans lequel Benguerel a montré que le flux d'air (airflow) de la nasale commence au plus tôt pendant la troisième consonne /f/ dans la séquence /rzfrV_n/.

Pour l'anticipation de la consonne sur les voyelles, les séquences /eC_n/, /eeC_n/, /eeeC_n/ ont été utilisées. La "C_n" représente une consonne nasale précédée par plusieurs voyelles non

nasales /e/. Les résultats montrent que, pour les séquences /eeC_n/ et /eeeC_n/ qui ont deux ou trois voyelles devant la consonne nasale, le mouvement vélaire commence à descendre plus tôt que pour la séquence /eC_n/ qui a une seule voyelle. Le geste d'abaissement du velum se produit dès que possible. Tout comme pour l'arrondissement labial, l'effet d'anticipation du geste vélaire semble n'avoir pas de limite à la frontière syllabique en français.

2.2.1.3 La coarticulation de voyelle à voyelle

L'existence de la coarticulation de voyelle à voyelle est toujours controversée en français. Certains chercheurs (Petursson et Bothorel, 1973 ; Marchal, 1985) montrent que la coarticulation de voyelle à voyelle n'est pas claire en français, mais les autres (Zerling, 1979 ; Vaxelaire et Sock, 1999, 2003) montrent des indices d'existence de coarticulation de voyelle à voyelle dans leurs études.

Petursson et Bothorel (1973) ont examiné la coarticulation du type V₁CV₂ en français en mesurant les transitions à partir de spectrogrammes sur un seul locuteur. Le corpus était constitué de consonnes occlusives /b d k g p t / et de voyelles / i e ε a y u o ã/. La différence de fréquence par rapport à la partie stable de la voyelle a été mesurée pour obtenir les transitions de fréquence. Les résultats montrent que, pour une séquence V₁CV₂, l'influence de V₂ sur la transition de V₁ et l'influence de V₁ sur la transition de V₂ en français sont moindres qu'en suédois et qu'en anglais américain dans l'étude d'Öhman (1966). Petursson et Bothorel trouvent qu'en français la préparation de la voyelle V₂ commencerait seulement pendant l'articulation de la consonne, mais non pas pendant l'articulation de la voyelle V₁, ni pendant la transition consonantique de la voyelle V₁. Mais ce résultat est obtenu par un seul sujet. Il est possible qu'il ne soit pas le cas général. La coarticulation est caractérisée par une grande variabilité individuelle et interindividuelle.

Marchal (1985) a étudié la coarticulation dans des groupes d'occlusive pour les séquences V₁C₁C₂V₂ par une méthode d'électropalatographie. Il a examiné séparément les consonnes en fonction de leurs articulations qui peuvent être produites par le même articulateur principal ou pas (Catford, 1977). Les consonnes occlusives produites avec les mêmes articulateurs comme /td/, /dt/, /kg/, /gk/ et les consonnes produites avec des articulateurs différents comme /tk/, /dg/, /dk/ et /gt/, ainsi que les géminées comme /dd/, /tt/, /kk/ et /gg/ sont analysées dans les séquences V₁C₁C₂V₂. Marchal a observé l'influence de la voyelle sur les consonnes, dans deux directions, l'influence de V₁ sur C₁ et l'influence de V₂ sur C₂. Mais l'influence directe de voyelle à voyelle n'a pas été observée. Les résultats ne sont pas en faveur de l'hypothèse

d'existence d'une coarticulation de voyelle à voyelle en français. Une possible explication tendrait à dire que, pour l'étude de la coarticulation vocalique, la méthode d'électropalatographie ne serait pas la plus adaptée. En ce qui concerne l'existence d'une continuité vocalique dans la parole, Marchal pense que si elle existe, elle est due à la persistance de la voyelle V_1 sur C_1 et de V_2 sur C_2 qui permet aux deux voyelles de se joindre en quelque sorte dans les consonnes. Il a indiqué que la coproduction ne concerne pas seulement les voyelles mais aussi les consonnes. "La coproduction vocalique coexiste avec la coproduction consonantique" (Marchal, 1985).

Zerling (1979) a étudié la coarticulation de voyelle à voyelle dans une séquence VCV et VCVCV en français par cinéradiographie. Les consonnes utilisées étaient /b/, /d/ et /g/ et les voyelles étaient /i, ε, a, ə, u/. Il a observé une coarticulation de voyelle à voyelle, ce qu'il a appelé "l'assimilation à distance" dans son étude. Les résultats montrent que dans certains cas, surtout pour une voyelle /ə/, l'anticipation peut s'effectuer à travers la consonne intervocalique. "Dans le groupes /əCV/, il n'est pas rare d'observer l'influence de V sur /ə/, lorsque la consonne occlusive le permet" (Zerling, 1979). Pour une séquence de /əCVCə/, il a observé que la première et la troisième voyelle, /ə/, sont influencées par la voyelle intermédiaire V, soit par anticipation soit par carryover.

Vaxelaire et Sock (1999, 2003) ont fait une étude de la coarticulation en français. L'influence d'une voyelle sur une autre voyelle est observée dans une séquence VCV. L'extension vers la gauche des caractéristiques labiales et linguales de la voyelle arrondie a été examinée à partir de données cinéradiographiques et de données acoustiques.

Les phrases ont été enregistrées par deux locuteurs français parlant avec deux rythmes de parole différents, rythme normal et rythme rapide. Le corpus était composé de séquences V_1CV_2 où la voyelle V_1 était une voyelle /i/ ou /a/, la voyelle V_2 était une voyelle arrondie /u/ ou /y/ et la consonne était /p/, /k/, /t/, /d/, /b/, /g/. La position et le déplacement horizontal et vertical des lèvres, de la mâchoire, de l'apex et du corps de la langue ont été mesurés à partir de la configuration de conduit vocal.

Les résultats montrent que, dans la séquence VCV, le geste labial de la voyelle arrondie /u/ en français peut s'étendre au-delà du début de la consonne intervocalique jusqu'aux dernières configurations de la voyelle non arrondie /a/. Le geste du corps de la langue peut également s'étendre à la première voyelle quand il n'est pas sollicité pour la production de la

consonne. Par exemple pour la séquence /atu/, le geste d'élévation du corps de la langue pour la voyelle arrondie /u/ peut s'étendre à travers la consonne intervocalique jusqu'à la première voyelle /a/. Enfin, ils ont observé que l'augmentation du débit d'élocution provoque l'accélération de l'activité articulatoire et la compression des durées acoustiques, mais qu'il ne modifie pas de manière significative l'étendue du geste d'anticipation.

En guise de conclusion, Vaxelaire et Sock proposent que, pour le geste du corps de la langue, l'anticipation de voyelle à voyelle existe quand ce geste n'est pas sollicité pour la consonne intervocalique. Dans le cas où le geste du corps de la langue est sollicité avec la consonne, par exemple dans la séquence /aku/, l'anticipation de voyelle à voyelle n'est pas observée. D'après Vaxelaire et Sock, l'étendue de l'anticipation du geste dépendrait en grande partie de facteurs tels que la nature du segment adjacent, le conflit des articulateurs dans la production de la séquence et des stratégies propres au locuteur.

2.2.2 La coarticulation en chinois

La phonétique et la phonologie modernes chinoises ont été développées à partir des années 1920 par deux grands phonéticiens chinois F. Liu et Y. R. Chao. Liu et Chao avaient fait leurs études à l'université de Paris et à l'université de Harvard. Ils ont apporté la phonétique et la phonologie occidentale à la phonologie traditionnelle chinoise, et ils ont utilisé au départ des données expérimentales. Le kymographe était l'outil principal pour les mesures des fréquences à l'époque. Ils ont écrit deux livres importants consacrés à la nouvelle étude du ton et de l'intonation en chinois : "*Experiments of four tones*" (Liu, 1924) et "*Study of modern Wu Dialects*" (Chao, 1928).

Les recherches en phonétique et phonologie en Chine commencent dès 1950 par Z-J. Wu. Bien que la coarticulation ait été largement étudiée dans les pays occidentaux, les travaux sur la coarticulation en chinois n'ont pas existé avant les années 80. A notre connaissance, l'articulation en chinois a été pour la première fois étudié par Wu (Wu, 1986, 1989). Il a étudié l'articulation et la coarticulation des voyelles et des consonnes dans la syllabe en chinois mandarin. Ensuite, la coarticulation entre les syllabes adjacentes, par exemple dans la séquence CVCV, a été analysée par des données acoustiques (Wu, 1991). Bao et Zheng ont utilisé pour l'étude de la coarticulation la méthode d'EPG sur la consonne ces cinq dernières années.

2.2.2.1 La coarticulation entre les syllabes

Wu (1991) a étudié la coarticulation sur des données acoustiques dans des séquences CVCV avec les consonnes occlusives. Le corpus a été constitué de combinaisons des consonnes /p/, /t/, /k/ et des voyelles /i/, /a/, /u/ en mandarin. Les effets de coarticulation entre les syllabes ont été analysés. Toutes les combinaisons de ces consonnes et de ces voyelles ont été prises y compris la combinaison /ki/ qui n'existe pourtant pas en chinois, (elle sert seulement pour la comparaison). Les effets sont mesurés par la transition de F_2 . Cinq points ont été analysés dans le spectrogramme : le début, le milieu et la fin de V_1 , le début et le milieu de V_2 . La transition T_2 représente l'amplitude de l'écart de fréquence par rapport à la partie stable de la voyelle. Les résultats montrent que, dans la première syllabe C_1V_1 , la transition de la consonne C_1 est toujours influencée par la deuxième syllabe C_2V_2 . La transition de la deuxième syllabe C_2V_2 est également influencée par la première syllabe. Pour ce qui est de la coarticulation dans la syllabe, Wu a indiqué que la dominance de la direction de coarticulation dans la syllabe CV (par exemple l'influence de V_1 sur C_1 et de C_1 sur V_1) est généralement différente selon la consonne, mais que, par contre, la différence entre les syllabes C_1V_1 et C_2V_2 n'est pas significative.

L'analyse de la coarticulation sur la transition de V_1C_2 à la frontière des deux syllabes est aussi analysée dans cette étude. Les résultats montrent que la transition de V_1 vers la consonne C_2 est essentiellement modifiée par la consonne C_2 . D'après Wu, dans la séquence / $C_1V_1C_2V_2$ / qui est composée de deux syllabes en chinois, une structure du type / V_1C_2 / à la frontière des syllabes ne forme pas une unité, donc la coarticulation peut être souvent bloquée par C_2 entre elles.

“A $C_1V_1C_2V_2$ function as two syllables in Standard Chinese with a phonological structure of CV as a solid unit each, so a structure of V_1C_2 does not form a unit and the coarticulation is usually blocked by C_2 between them.” (Wu, 1991)

Les résultats montrent que les syllabes s'influencent les unes les autres en chinois. Cela va dans le sens de l'hypothèse de Kozhevnikov et Chistovich (1965) pour qui la syllabe est considérée comme l'unité de base de l'articulation. Par contre, ces observations de la langue chinoise infirment l'hypothèse d'Öhman selon laquelle il existerait une base vocalique continue dans la parole.

2.2.2.2 La coarticulation dans la syllabe

Bao (2001) a étudié des données d'EPG du chinois avec un corpus de 62 syllabes CV qui concernent cinq voyelles /a, u, i, ɿ, ʅ/ (/ɿ/ existe seulement en contexte de /s ɿ/ et /ts ɿ/, et / ʅ / existe seulement en contexte de /ʃ ʅ/et /tʃ ʅ/) et 21 consonnes. Les contacts linguo-palataux des consonnes ont été examinés. Les résultats révèlent que parmi les cinq voyelles, /a/ et /u/ ne génèrent pas de contacts dans la zone de mesure, alors que /i/ en génère un important. Chaque groupe de consonnes a ses propres contacts. Les voyelles /a, i, u/ ont des effets de coarticulation différents sur les consonnes, et ces effets existent pendant l'articulation de la consonne dans une syllabe CV. Normalement les consonnes /b, p, m/ n'impliquent pas de contacts dans la zone concernée. Elles peuvent cependant être influencées par la voyelle /i/ qui suit. Les autres consonnes subissent aussi l'effet d'anticipation de la voyelle /i/ et les contacts deviennent plus antérieurs. Parmi eux, la consonne /n/ montre un large effet de palatalisation, par exemple pour /ni/. La zone de contact des consonnes occlusives /g, k/ est plus antérieure quand ces consonnes sont suivies d'une diphtongue /ei/. Mais suivi d'une voyelle /u/, les contacts des consonnes occlusives sont davantage postérieurs. Ces effets peuvent être expliqués par l'anticipation de la voyelle suivante sur la consonne dans une syllabe.

Cette étude a permis de recueillir des informations sur les lieux d'articulation des consonnes en chinois et sur la coarticulation dans la syllabe CV, informations qui constituent une base pour toute étude ultérieure.

2.2.2.3 La coarticulation de N_1C_2 en chinois

Zheng et Bao (2003) ont étudié la coarticulation du type N_1C_2 où N_1 est une consonne nasale. En chinois, il y a trois consonnes nasales /m/, /n/ et /ŋ/. Parmi elles, seulement deux peuvent apparaître en position initiale dans une syllabe, /m/ et /n/, et seulement deux peuvent apparaître en position finale, /n/ et /ŋ/. Zheng et Bao (2003) ont examiné de quelle façon la nasale N_1 , coda de la première syllabe, peut être influencée par la consonne C_2 , attaque de la syllabe suivante. C'est ce qu'on appelle "la coarticulation N_1C_2 " dans une structure / $C_1V_1N_1C_2V_2$ /.

Zheng et Bao ont utilisé la technique de l'EPG pour analyser cette coarticulation N_1C_2 . Les résultats ont été mis en relation avec les données acoustiques. Les effets sur la nasale N_1 de la consonne C_2 et de la voyelle V_2 ont été analysés. Les auteurs ont montré que, dans la

séquence $/C_1V_1N_1C_2V_2/$, les gestes de fermeture (closing) et une partie de fermeture consonantique (closed) de C_2 sont accomplis pendant le geste d'abaissement du vélum de N_1 . Zheng et Bao ont appelé ce phénomène chevauchement temporel non-linéaire (*“temporal non-linear overlap”*).

En ce qui concerne l'influence de la voyelle V_2 sur N_1 , Zheng et Bao ont indiqué que la coarticulation de N_1C_2 n'est pas influencée par V_2 en général. Cette étude de Zheng et Bao montre que les locuteurs chinois semblent commencer chaque syllabe par une consonne initiale dans la syllabe précédente. L'effet de la voyelle de la nouvelle syllabe sur la syllabe précédente est limité. Cet effet est fortement influencé par la consonne initiale de la nouvelle syllabe. Ces résultats confirment ainsi l'hypothèse de Kozhevnikov et Chistovich (1965).

Chapitre 3 : Bilan : Enjeux et démarche méthodologique

Introduction

Après avoir présenté dans les 2 chapitres précédents un état de l'art des principales théories de la coarticulation et quelques aspects pertinents de la phonologie du français et du chinois pour notre étude, nous présentons dans ce chapitre les enjeux de nos travaux sur la modélisation de la planification de la production de parole. Nous évoquerons quelques questions, essentielles à nos yeux, sur la planification des séquences de parole, ayant trait à la nature de la planification motrice et à son interaction avec les caractéristiques linguistiques. Nous verrons aussi comment nous avons décidé d'aborder ces questions, d'une part par le biais d'une étude expérimentale sur les locuteurs natifs du français et du chinois, et, d'autre part en évaluant quantitativement les hypothèses de planification en élaborant une modélisation du contrôle et en comparant les simulations obtenues aux données.

3.1 Les enjeux de notre modélisation de planification

Dans notre étude, nous favorisons l'hypothèse selon laquelle la coarticulation est planifiée. La planification de la parole affecte directement les commandes motrices dans le Système Nerveux Central du locuteur. Il a été suggéré qu'elle est le résultat d'une stratégie de contrôle de niveau supérieur (Perrier, 2004). Les questions essentielles de la planification sont de savoir dans quel espace le locuteur planifie et donc de quelle nature est la tâche de parole, motrice, articulatoire, ou perceptive. Quel outil pouvons-nous utiliser pour modéliser la planification ? Quel critère nous pouvons appliquer pour un contrôle optimal ? Et quelle interaction linguistique la planification concerne-t-elle ?

3.1.1 Planifier dans quel espace ? Avec quelle complexité ?

La tâche de planification en production de parole consiste d'abord en une transformation d'une séquence d'entités phonologiques abstraites en séquences d'objectifs dans un espace de contrôle. Un tel espace pourrait être l'espace articulatoire ou l'espace acoustique ou un espace perceptif ou encore un espace multimodal associant ces différents espaces potentiels. La spécification de cet espace renvoie à la question fondamentale de la représentation de la tâche de parole dans un espace physique. Pour cela, il est intéressant de se placer dans une perspective d'interaction entre la production de la parole et sa perception : ce qui est perçu par l'auditeur, serait ainsi ce que le locuteur cherche à produire, et inversement, les intentions du

locuteur sont ce que cherche l'auditeur dans le flux de la parole. C'est le concept de monnaie d'échange, développé en particulier par Goldstein et Fowler (2003). Ainsi les théories de perception offrent des pistes pour comprendre les objectifs de la production. L'hypothèse de la théorie Motrice de la perception de la parole (Lieberman et Mattingly, 1985) propose que ces objectifs soient caractérisés par des invariants moteurs, correspondant à une intentionnalité du locuteur. Cet invariant moteur est le geste articulatoire que le modèle de la dynamique de la tâche a modélisé comme le mouvement vers un attracteur dynamique. La théorie de perception directe de la parole (Fowler, 1986) a proposé que l'invariant physique existe sous la forme de conduit vocal. Autrement dit, l'objectif de la production de la parole serait articulatoire et mesurable par la géométrie du conduit vocal. Stevens (1972, 1978) a proposé dans sa théorie quantique que l'objectif associé à la production d'un phonème donné se trouve dans le signal acoustique et plus spécifiquement dans les caractéristiques spectrales du signal. Guenther (1998, 2006) a proposé de caractériser les cibles de son modèle DIVA par des régions multimodales associées à des objectifs auditifs et orosensoriels (cf. le chapitre 1 pour un développement plus général sur le modèle DIVA). Schwartz (2002) a décrit dans la théorie de perception pour le contrôle de l'action (PACT) que la perception de la parole est l'ensemble des processus perceptifs multimodaux (auditifs, visuels...) qui permettent de récupérer et de contraindre les cibles des gestes vocaux. L'absence d'une modalité peut être compensée par une autre. Perrier (2005b) a repris l'hypothèse d'une représentation multimodale selon laquelle la représentation de la parole n'est pas purement acoustique ou purement articulatoire, et propose aussi qu'en conditions normales les deux composants coexistent. Cependant, Perrier (2005b) considère que la composante acoustique est prioritaire dans la spécification des objectifs et leur réalisation. L'objectif premier est donc de produire le son.

Le deuxième volet de la planification motrice consiste à trouver les séquences motrices permettant les réalisations objectives. La difficulté de cette étape vient du fait que la relation entre commandes motrices et caractéristiques de sortie n'est pas biunivoque. Pour une position cible dans l'espace des tâches, il existe donc plusieurs patrons moteurs qui lui sont associés. Pour résoudre cette difficulté et contraindre les solutions, la démarche classique consiste à augmenter les contraintes par l'évaluation d'un coût. Il s'agit alors de l'élaboration de stratégies motrices optimales. La planification sur la base de telles stratégies fait appel à une représentation générale de relation entre entrées et sorties, en d'autres termes à un modèle interne direct (Jordan, 1990, 1992 ; Kawato, 1990). Dans ce contexte, une autre question

fondamentale est celle de la complexité du modèle interne : jusqu'à quel point et avec quelles précisions ce modèle doit-il être capable de rendre compte de la physique du système. Certains auteurs ont proposé qu'un modèle interne statique qui décrit une relation entre le statut statique des positions articulatoires du conduit vocal et les caractéristiques spectrales pourrait être suffisant. (Savariaux *et al.*, 1995 ; Perkell *et al.*, 2000). Mais Kawato (1990) a proposé un modèle interne décrivant le système dans toute sa complexité, aérodynamique et dynamique. Sur cette question pour la parole, Perrier (2006) a proposé que malgré la complexité de la tâche de la production de la parole, un modèle interne simple pourrait convenir pour le contrôle de la production de la parole qui peut être associé par des boucles feedback de bas niveau (ne modifie pas les commandes envoyées par le Système Nerveux Central). Il a montré que la réalisation de trajectoires complexes pouvait être obtenue avec un modèle interne statique, décrivant les relations entre activation musculaire et position de la langue. La complexité des trajectoires serait alors due aux caractéristiques biomécaniques de la langue.

3.1.2 Le modèle interne et la planification

De nombreux modèles de contrôle du mouvement humain, exploitant le concept de modèle interne, ont été élaborés. (Jordan, 1990 ; Jordan et Rumelhart, 1992 ; Kawato *et al.*, 1990). Dans la production de la parole, ce concept est exploité (Bailly *et al.*, 1991 ; Guenther, 1995, 1998 ; Perkell *et al.*, 2000 ; Perrier, 2005), afin de réaliser de façon optimale des buts dans un domaine acoustique, perceptif et/ou articulatoire. Mais la façon d'intégrer le modèle interne dans la chaîne de contrôle est encore très variable selon les travaux.

Kawato et ses collègues (Kawato *et al.*, 1987 ; Kawato *et al.*, 1990) ont proposé pour la modélisation du contrôle du mouvement humain d'exploiter deux modèles internes : un modèle direct simulant le système moteur lui-même et rendant donc compte de manière fonctionnelle des relations entre commandes motrices et variables de sortie, et un modèle inverse permettant d'inférer les commandes motrices à partir de la spécification des objectifs à atteindre dans l'espace distal (l'espace de description de la spécification de la tâche). Cela permet d'élaborer un système de contrôle du type feedforward/feedback. Le contrôleur feedforward exploite le modèle interne inverse pour proposer des commandes motrices aptes à la réalisation de la tâche proposée. Le contrôleur feedback prend en compte les sorties que prédit le modèle direct (on parle de copie d'efférence) à partir des commandes motrices. Il calcule l'erreur entre ces valeurs de sorties prédites et celles spécifiées dans l'espace distal. Si cette erreur dépasse un seuil pré-établi, le contrôleur feedback calcule (avec le jacobien

inverse du modèle direct) les variations à apporter aux commandes motrices pour réduire cette erreur. Dans le cas où la tâche motrice le permet, le contrôleur feedback peut aussi prendre en compte les informations sur les variables de sortie effectivement atteintes par le système moteur proprement dit. Cela implique cependant que la durée d'exécution de la tâche soit suffisamment longue pour tolérer les retards associés à la propagation des informations efférentes (feedback) vers le cortex moteur et à leur traitement. Un certain nombre de publications ont montré que le traitement cortical de ce type de feedback, que nous qualifierons de haut niveau, implique des durées de traitement supérieures à 50 ms (voir par exemple pour le traitement du feedback auditif, Siegel et Pick, 1974). La durée des gestes de la parole pouvant être aussi brève que quelques dizaines de millisecondes, nous n'intégrerons pas cette possibilité dans notre analyse.

Le modèle inverse peut être élaboré de manières différentes. En ce qui concerne la production de la parole, il vise à retrouver les commandes motrices à partir de la spécification de la tâche (voir ci-dessus pour les discussions concernant la nature de cette spécification). Ce problème étant many-to-one, nous l'avons dit, ce modèle inverse n'est pas unique. Sa nature dépendra du critère utilisé dans la phase d'optimisation permettant de restreindre l'espace des solutions. Cette optimisation peut s'opérer dans l'espace des commandes (Jordan, 1989), dans celui des variables de sorties (Jordan et Rumelhart, 1992) ou dans un espace intermédiaire entre ces deux espaces (Kawato *et al.*, 1990).

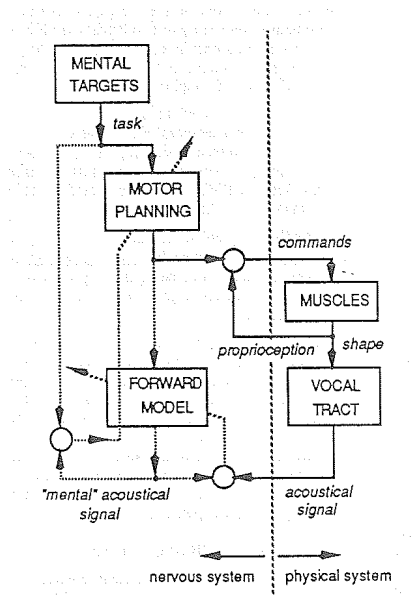


Figure 3.1 La modélisation de contrôle moteur (Laboissière, 1990)

Laboissière *et al.* (1990) ont proposé un schéma cadre pour décrire les processus d'apprentissage de ce modèle inverse. Il est donné à la figure 3.1. Les lignes pleines représentent des commandes motrices et les signaux physiques du processus de production de la parole. Les cibles mentales sont des représentations abstraites des séquences (phonèmes). Elles spécifient la tâche phonologique pour la planification motrice ("Motor planning"). Cette planification motrice est en charge de la génération des commandes motrices en vue de produire une sortie acoustique porteuse d'information linguistique. Les lignes pointillées représentent les informations pendant la phase d'apprentissage. Un modèle interne direct ("Forward model") est employé pour décrire la relation entre les commandes motrices et les sorties acoustiques.

Dans une première phase, le cerveau apprend le modèle interne direct en minimisant l'erreur globale entre les variables de sortie et les prédictions du modèle. Nous l'avons dit, la complexité de ce modèle pourra varier selon les hypothèses faites sur la nature des variables de sorties prises en compte. Ce modèle direct va ensuite être exploité pour apprendre le modèle inverse ("Motor planning"). L'inférence des commandes motrices aptes à la réalisation d'objectifs passe par une phase de simulation interne exploitant le modèle interne direct. Des valeurs initiales sont proposées pour les commandes motrices. Leur impact sur les variables de sortie est alors prédit par le modèle direct. La connaissance des variables motrices et de la copie d'efférence permet d'estimer le coût à optimiser. On utilise alors des connaissances sur la sensibilité de ce coût à la variation locale des commandes motrices (Jacobien inverse) pour proposer une variation des commandes motrices qui permettra de réduire ce coût. Les conséquences de ces nouvelles commandes motrices sur le coût sont à leur tour évaluées, et ainsi de suite jusqu'à ce qu'on obtienne le coût minimum. Les commandes motrices ainsi trouvées sont celles que fournissent le modèle inverse pour la génération de la séquence de parole souhaitée ("Motor planning"). Dans cette représentation, le contrôleur feedback sert à élaborer le modèle inverse. Il n'est pas exploité au cours de la production de la parole proprement dit.

Laboissière *et al.* (1990) ont proposé d'exploiter un modèle direct statique associant commandes motrices et formants, et leur modèle inverse minimisait un coût de lissage global sur les trajectoires des 3 premiers formants. On peut alors parler de modèle inverse cinématique. Cette proposition rejoint celle de Jordan et Rumelhart (1992) pour les mouvements de pointage séquentiel du bras. Kawato *et al.* (1990) ont proposé que le modèle direct rende compte des caractéristiques dynamiques du système moteur et que le coût à

minimiser pour l'élaboration du modèle direct est le couple global mis en jeu dans la séquence gestuelle. On peut alors parler de modèle inverse dynamique. Pour Jordan (1989) le coût était directement dans l'espace moteur. Il visait à minimiser le chemin parcouru dans cet espace au cours de la séquence motrice. Il ne faisait donc pas appel à un modèle direct et son modèle inverse peut être qualifié de statique. Perrier *et al.* (2005) se sont inspirés de ces propositions pour la production de la parole. Mais nous y reviendrons plus amplement dans ce manuscrit. Toutes ces propositions ne font pas appel au contrôleur feedback au cours de la production de la parole.

Le modèle DIVA proposé par Guenther (1995, 1998, 2006) se différencie un peu de ces propositions sous deux aspects. Le contrôleur feedback exploitant la copie d'efférence est un élément clé de la stabilité du contrôle. Il permet de corriger les imprécisions des commandes motrices proposées à partir de la spécification de la tâche linguistique. Le lien entre la spécification de la tâche linguistique et les commandes motrices repose par ailleurs essentiellement sur un assocateur (comparable, nous l'avons dit au chapitre 1, au syllabaire mental de Levelt, 1994) et non sur un modèle inverse.

Pour l'apprentissage des différents modèles internes, Jordan (1988, 1990, 1991) a proposé une méthode d'apprentissage supervisé s'appuyant sur un réseau de neurones multicouche et s'appuyant sur l'algorithme de rétropropagation du gradient. (Rumelhart, Hinton et Williams, 1986). Cette méthode effectue essentiellement une descente du gradient du critère à optimiser (erreur de prédiction pour le modèle direct et coût pour le modèle inverse) dans l'espace distal. En ce qui concerne le modèle interne direct, l'enjeu de la modélisation est sa capacité à proposer une généralisation à partir d'un nombre suffisant de données d'apprentissage. Cette capacité à généraliser est essentielle puis que le modèle interne a pour objectif de prédire des configurations qui n'auraient pas encore été produites. Ceci peut être particulièrement utile pour gérer les perturbations.

3.1.3 Critère de planification

Comme nous l'avons dit, pour retrouver les commandes motrices à partir de sorties désirées, il faut des optimisations utilisant certains critères qui permettent la régularisation du problème many-to-one.

Nelson (1983) a proposé des critères du type contrôle moteur que l'on peut résumer par un objectif global de minimisation de « l'effort » du locuteur. Cet effort pourrait, selon Nelson, être la distance, la durée, la force, le jerk, ou la vitesse du mouvement. Pour la parole, la

planification consisterait alors à minimiser la somme des «efforts» du locuteur mis en jeu au cours de la production d'une séquence.

La tâche de la parole est, nous l'avons montré ci-dessus, une tâche qui n'est pas spécifiée de manière unique dans l'espace physique. La variabilité associée à la réalisation d'une entité linguistique est à la fois liée aux phénomènes de coarticulation évoqués dans notre premier chapitre et au style d'élocution, qui peut être fait à un débit rapide ou lent. Ces variabilités observées dans la production de la parole ont été formalisées par Lindblom (1990) avec les concepts de négociation entre auditeur et locuteur, et la parole hyper-articulée ou hypo-articulée. Le principe de base des propositions de Lindblom est que la qualité de la production de la parole serait le résultat d'une négociation entre les exigences de minimisation de l'effort du locuteur et celle de la production d'un signal identifiable par l'auditeur. Une parole hyper articulée résultera d'une priorité donnée à la réalisation des objectifs perceptifs. Une parole hypo articulée sera la conséquence d'un avantage donné à la minimisation de l'effort du locuteur. Dans la lignée de ces propositions, pour le contrôle de production de la parole, le critère à minimiser possède une composante perceptive, en plus de la composante motrice. (cf. Bailly *et al.*, 1990). La question des points respectifs de ces deux composantes dans la définition du critère à optimiser est elle aussi posée. En jouant sur ces poids, on pourra jouer sur la qualité (hyper/hypo) de la parole produite.

3.1.4 Interaction linguistique de planification

La planification de la parole qui peut intégrer une composante perceptive, est probablement spécifique par rapport à la planification d'autres mouvements humains parce qu'elle peut être contrainte par des facteurs linguistiques liés à la structure phonologique de langue. La structure phonologique de langue, avec en particulier les caractéristiques des unités phonologiques de base, peut notamment influencer la longueur de séquences pour lesquelles le critère de planification est optimisé. Une séquence plus ou moins longue pourrait être ainsi prise en compte pour la planification. Le chapitre 2 a présenté différentes hypothèses sur la nature des unités phonologiques. Elle pourrait être une syllabe CV, une séquence VCV, un mot ou même une unité prosodique (cf. Le schéma de Jun et Fougeron 2002 pour unité prosodique).

A la suite des propositions de Kozhevnikov et Chistovich (1965), on pourrait proposer que la syllabe soit l'unité de planification dans le contrôle moteur de la parole. Kent et Minifie (1977) ont fait des propositions et ont suggéré que la planification de contrôle moteur

est relative à des intervalles spécifiques, délimités par les voyelles. Pour eux, les consonnes seraient ainsi planifiées avec les voyelles dans la syllabe comme une entité. Wood (1997) est aussi allé dans ce sens lors de l'analyse de données cinéradiographiques en suédois et en Bulgare (Wood, 1991, 1996).

Si on adhère au modèle élaboré par Öhman, on choisira la séquence vocalique VV comme séquence de base dans la planification. Le modèle hybride de Perkell (1992) et le modèle MEM d'Abry (1991) inciteront plutôt à prendre des séquences encore plus grandes comme entité temporelle de planification.

Le choix de la nature et le type de séquence dans la planification sont probablement liés à la phonologie de chaque langue. Notre travail expérimental et de modélisation sur le français et le chinois a pour objectif d'étudier de manière quantitative cette hypothèse. C'est dans ce but que nous avons élaboré une démarche méthodologique dont nous décrirons les grandes lignes ci-dessous.

3.2 Méthodologie

3.2.1 Une étude expérimentale en français et en chinois

En ce qui concerne la question de la planification d'interaction linguistique avec le niveau linguistique, nous nous demandons si le choix de l'unité de planification est le même pour le français et le chinois. Nous avons présenté dans le chapitre 2 un certain nombre d'éléments suggérant que pour le chinois la syllabe serait une unité fortement structurante, et que le statut de la syllabe en français n'aurait pas la même importance au niveau phonologique.

Sur le plan expérimental, nous allons analyser des données articulatoires pour observer la coarticulation d'anticipation des locuteurs français et des locuteurs chinois. Cette analyse portera spécifiquement sur les positions cibles atteintes par la langue pour chacun des sous éléments de la séquence de parole. Cela sera fait en analysant les positions de 4 capteurs placés sur le contour de la langue dans le plan sagittal de la tête avec le système EMMA.

Pour une séquence V_1CV_2 , nous allons analyser l'anticipation de la voyelle V_2 sur la voyelle précédente V_1 (anticipation non adjacente et hors cadre syllabique) et l'anticipation de la voyelle V_2 sur la consonne C (anticipation adjacente à l'intérieur du cadre syllabique). L'existence de l'anticipation de V_2 sur V_1 reflète une stratégie d'anticipation sur une unité plus longue que la syllabe. En revanche, sans l'anticipation de V_2 sur V_1 , la planification sera limitée à la syllabe CV_2 .

3.2.2 Une modélisation du contrôle moteur

3.2.2.1 La méthodologie d'une modélisation du contrôle moteur

Nous proposons une nouvelle modélisation du contrôle moteur dans cette étude basé sur l'hypothèse du point d'équilibre (Feldman, 1986), sur l'hypothèse de représentation interne (Kawato, 1990) et sur l'hypothèse de contrôle optimal (Nelson, 1983). L'hypothèse de base du point d'équilibre propose que le mouvement soit recruté par des changements des variables neurales de commande qui déplacent le point d'équilibre du système moteur (Feldman, 1986). Le rôle d'hypothèse du point d'équilibre repris par Perrier *et al.* dans la parole consiste à envisager le contrôle d'un modèle articulatoire dans lequel chaque articulateur, caractérisé par son propre mécanisme de muscle, se déplace vers une cible. Cette hypothèse a décrit comment la trajectoire articulatoire désirée peut atteindre sa cible par chaque articulateur avec sa propriété mécanique et comment la commande motrice peut être liée avec la phonologie de la langue. En définissant le mouvement par le déplacement d'un statut vers un autre statut, l'hypothèse du point d'équilibre permet de produire une trajectoire articulatoire continue à partir d'une succession de commandes discrètes qui peuvent être liées à la séquence phonologique discrète.

Nous faisons l'hypothèse que la tâche de parole est prioritairement spécifiée dans un domaine acoustico-perceptif (Perrier, 2005b) et l'hypothèse de l'existence de représentations internes simples statiques dans le système du contrôle de la parole (Perrier, 2006). Par conséquence, notre modélisation fait appel à une description de système moteur plus simple que celui de Jordan (1992) ou Kawado *et al.* (1990). Nous faisons donc l'hypothèse de la parcimonie pour la structure du système de contrôle de la production de la parole. En d'autres termes, nous privilégions les solutions simples et testons leurs capacités à rendre compte de la réalité. Le modèle interne que nous allons exploiter n'intègre donc pas toute la complexité du système de production de la parole. Nous ne spécifions pas la trajectoire pas à pas, mais nous définirons des objectifs successifs, c'est à dire des cibles dans l'espace physique qui sont reliées aux phonèmes. La planification se fait sans le jacobien inverse du modèle interne direct, mais en exploitant le modèle interne direct uniquement pour évaluer le coût perceptif associé à des commandes motrices données. L'optimisation est définie dans le but de bouger le moins possible dans l'espace des commandes motrices afin de minimiser « l'effort » du locuteur. Les contraintes perceptives sont définies par rapport à des régions perceptives cibles de chaque phonème. Notre modélisation n'intègre pas de contrôleur feedback que ce soit par copie d'efférence ou par feedback réel provenant du système physique (il ne serait donc pas

apte à détecter une perturbation en terme du type bite-block). Par contre le modèle de contrôle de bas niveau, générant des forces musculaires à partir de commandes envoyées par le SNC prend compte un feedback de faible latence qui permet la correction des forces sans changer les commandes motrices et donc sans réorganisation de la stratégie.

Ce contrôle moteur est appliqué à un modèle biomécanique de la langue (Payan et Perrier, 1997 ; Perrier *et al.*, 2003). Nous allons utiliser cette modélisation pour tester différentes stratégies de planification et étudier l'évolution temporelle des commandes motrices entre les cibles. Pour cela nous utilisons aussi différentes trajectoires dans l'espace moteur entre les cibles successives inférées lors de la phase d'optimisation/planification. C'est ce que nous appelons les stratégies d'exécution du mouvement.

3.2.2.2 Hypothèse de planification et d'exécution du mouvement

Trois hypothèses de planification et deux hypothèses d'exécution du mouvement seront envisagées dans notre étude. Toutes les hypothèses reposent sur le concept de cible, qui veut que les gestes de la parole soient dirigés vers des objectifs articulatoires, acoustiques ou perceptifs associés spécifiquement à chacun des éléments de la chaîne phonétique, le phonème. Les trajectoires suivies par les articulateurs entre deux cibles seraient le résultat de la façon dont le système moteur cherche à atteindre les cibles successives. C'est la spécification motrice de planification de ces cibles et l'évolution temporelle des commandes motrices entre les cibles qui différencieront les hypothèses qui seront évaluées.

A. Hypothèse de planification

L'hypothèse de planification vise à minimiser le chemin parcouru entre les cibles dans l'espace des commandes motrices.

a. Hypothèse de planification de voisinage

L'hypothèse de planification de voisinage prend en compte tous les phonèmes de manière équivalente. C'est-à-dire qu'il n'y a pas de différence de statut entre les consonnes et les voyelles : tous ces phonèmes sont pris en compte de la même façon et avec la même importance. Toutes les commandes varient systématiquement de la même manière entre les commandes cibles. Pour une séquence V_1CV_2 , la planification est donc définie par le rapprochement global de toutes les cibles de la séquence V_1-C-V_2 .

b. Hypothèse de planification syllabique

L'hypothèse de planification syllabique est une spécialisation de l'hypothèse de voisinage

à la syllabe. Pour une séquence V_1CV_2 , elle donne un statut spécifique à la syllabe CV_2 .

c. Hypothèse de planification inspirée du modèle d'Öhman (1967)

L'hypothèse de planification selon Öhman est implantée également par un rapprochement de voisinage au sens de la séquence vocalique V_1V_2 .

B. Hypothèse d'exécution du mouvement

Suite à l'hypothèse de planification, nous avons toutes les cibles planifiées pour produire une séquence VCV. Mais la production de séquences est également influencée par les caractéristiques cinématiques sur le système du mouvement. Pour savoir comment générer les mouvements entre cibles, nous testerons deux types d'hypothèse sur l'exécution du mouvement.

a Hypothèse d'exécution séquentielle

L'hypothèse d'exécution séquentielle ne fait pas de différence entre consonnes et voyelles dans une séquence V_1CV_2 , et les commandes varient systématiquement de la même manière entre les valeurs cibles ; Le temps de transition entre toutes les positions de cible d'une séquence est égal à 40ms et le temps de tenue des cibles à 100ms.

b Hypothèse d'exécution selon Öhman

L'évolution des commandes motrices ne se fait donc pas séquentiellement pour une séquence V_1CV_2 . Elle donne un statut particulier aux transitions de voyelle à voyelle, les consonnes étant conçues alors comme des perturbations localisées dans le temps de cette base vocalique. Le temps de transition vocalique est défini à 180ms. Les temps des transitions vers la cible consonantique sont de 25ms. En comparaison avec l'hypothèse d'exécution séquentielle, les temps de tenue pour toutes les cibles restent à 100ms.

Deuxième partie : Méthodologie

Chapitre 4 : Démarche pour notre modélisation du contrôle moteur de la parole

Introduction

Ce chapitre a pour but de présenter notre modélisation du contrôle moteur de la parole. Cette modélisation sera exploitée pour tester différentes hypothèses sur la planification et l'exécution de séquences VCV de parole (Chapitre 8). L'objectif de ce chapitre est de décrire la méthode utilisée pour la modélisation à partir des commandes motrices jusqu'aux signaux acoustiques avec un modèle biomécanique de la langue, et pour l'inversion d'un espace acoustico-perceptif aux commandes motrices avec les critères optimaux grâce à un modèle interne (Jordan 1990,1992; Kawato *et al.*, 1990; Hirayama *et al.*, 1992;)

Après une présentation du modèle biomécanique de la langue (Payan et Perrier, 1997), nous présentons successivement :

- La génération de la base de simulations pour les formes de la langue et pour les caractéristiques acoustiques.
- Un modèle interne qui décrit les relations entre les commandes motrices et les caractéristiques acoustiques. Des fonctions radiales de bases ("*Radial Basis Functions*") seront employées pour élaborer ce modèle.
- Une inversion optimale des commandes motrices à partir de régions de cibles du phonème dans un espace acoustico-perceptif.

4.1 Modèle biomécanique

Le modèle biomécanique de la langue consiste en une structure à éléments finis représentant la projection de la langue dans le plan sagittal de la tête (Payan et Perrier, 1997 ; Perrier *et al.*, 2003). Le mécanisme de génération et de contrôle de la force musculaire est inspiré de la théorie du point d'équilibre de Feldman (1986). Les sept principaux muscles déterminant la forme de la langue dans le plan sagittal sont modélisés par des macro-fibres

généralisant des forces en certains nœuds de la structure. Chaque muscle est représenté dans la structure à éléments finis par un sous-ensemble spécifique d'éléments dont les propriétés élastiques varient avec le niveau d'activation du muscle. Les forces musculaires sont appliquées en certains nœuds par des macro-fibres dont les directions principales ont été ajustées pour décrire au mieux les données anatomiques sur les orientations majoritaires des fibres musculaires. Les muscles implémentés sont le Génio-glosse Postérieur, le Génio-glosse Antérieur, le Stylo-glosse, l'Hyoglosse, le Verticalis, le Longitudinalis inférieur et le Longitudinalis supérieur.

La force générée par chaque muscle est contrôlée via un paramètre de commande λ , propre à chacun des muscles et assimilable à une longueur conformément aux propositions de la théorie du point d'équilibre de Feldman (1986). Ce paramètre est appelé "seuil de recrutement musculaire". Lorsque la longueur du muscle est plus petite que (ou égale à) ce seuil de recrutement associé, aucune force n'est générée par ce muscle. Si le muscle est de longueur plus grande que son seuil de recrutement, la force augmente de façon exponentielle avec l'écart entre cette longueur et le seuil λ . La relation entre commandes musculaires et forces musculaires est donc typiquement non linéaire, avec en particulier un effet quantique pour chaque muscle de part et d'autre de la valeur λ spécifiant son seuil de recrutement. Le modèle d'Yong des éléments de la structure à éléments finis correspond au muscle actif qui augmente linéairement avec l'activation.

Le modèle intègre les contacts palataux. Ce modèle a été jusqu'à aujourd'hui essentiellement exploité pour tester l'influence des propriétés physiques des articulateurs sur les trajectoires de la langue au cours de la production de la parole (Perrier *et al.*, 2000 ; Perrier *et al.*, 2003).

La figure 4.1 présente quelques exemples de déformation de la langue générés grâce à ce modèle :

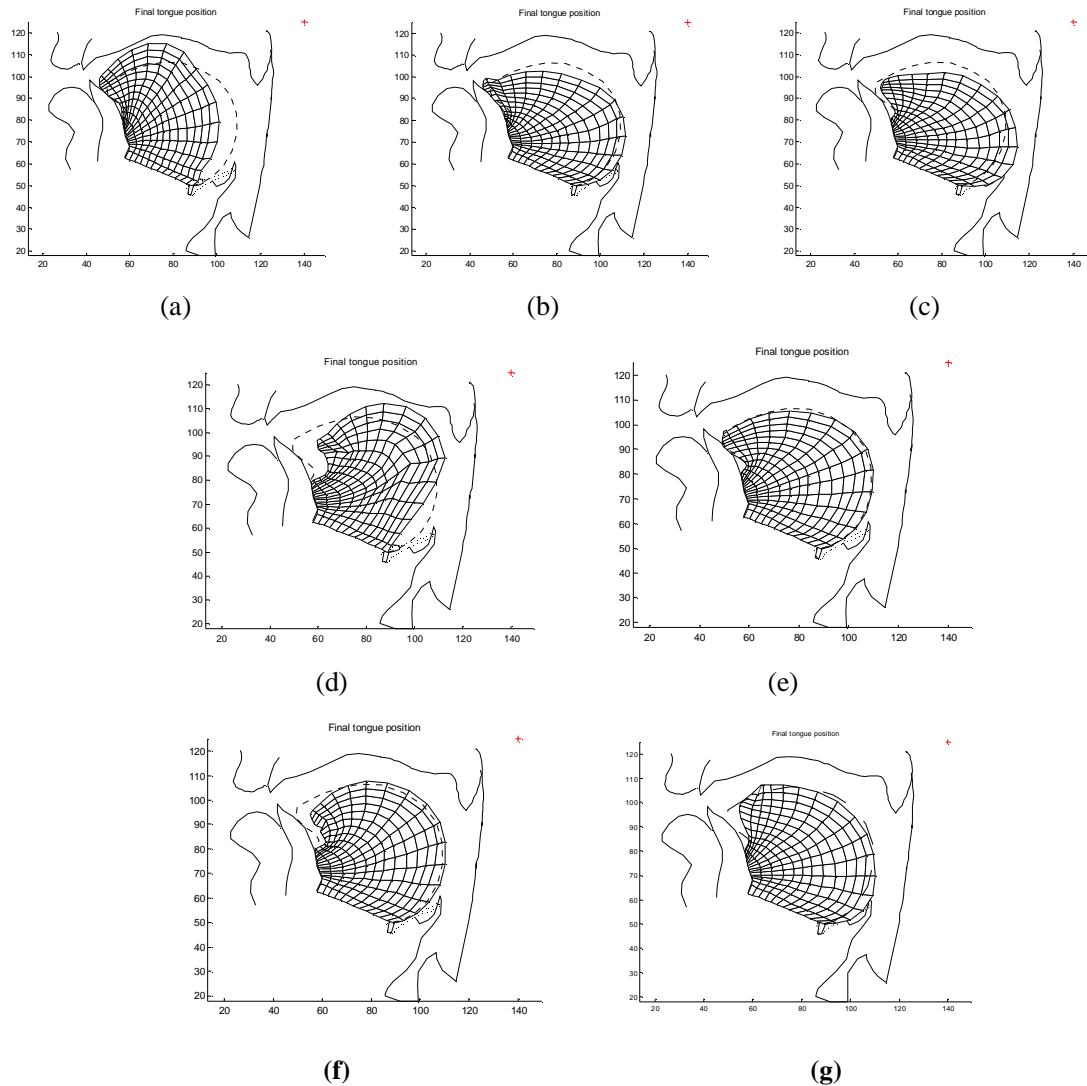


Figure 4.1 : Exemples de configurations obtenues avec le modèle de Payan et Perrier (Version Perrier *et al.*, 2003): activation du muscle Génio-glosse Postérieur (a), activation du muscle Génio-glosse Antérieur (b), activation du muscle Hyoglosse (c), activation du muscle Styloglosse (d), activation du muscle Verticalis (e), activation du muscle Longitudinalis inférieur (f) et activation du muscle Longitudinalis supérieur (g).

Le trait en pointillés représente le contour de la langue dans sa position de repos.

4.2 Constitution de la base de simulation

Le contrôle de production de la parole concerne un concept de modèle interne (Jordan 1990, 1992; Kawato *et al.* 1990; Hirayama *et al.* 1992). Le modèle interne correspond à une représentation que le système nerveux central aurait des relations entre des commandes motrices et les caractéristiques acoustiques, et qu'il exploiterait pour élaborer les séquences motrices permettant la production d'une séquence de parole.

Pour élaborer ce modèle interne, nous avons préalablement généré un ensemble de simulations avec le modèle biomécanique, à partir duquel, en exploitant des techniques d'apprentissage classique, il devient possible d'élaborer une description générale des relations entre l'espace de commande et l'espace de sortie.

4.2.1 Génération des formes de langue par la Méthode de Monte-Carlo

Les simulations ont été générées grâce à un échantillonnage de l'espace des variables de commande, les lambdas, fondé sur une méthode de Monte-Carlo intégrant l'hypothèse d'une distribution uniforme des paramètres de commande autour de la forme neutre de la langue (cf. Khaligi, 1999 ; Perrier *et al.*, 2000). La dynamique d'échantillonnage de chaque grand muscle était entre -20mm et 20mm. Pour le muscle longitudinalis inférieur et le muscle verticalis, elle était entre -10mm et 10mm. Le muscle longitudinalis supérieur n'a pas été retenu pour cause d'instabilité numérique trop fréquente.

La base des simulations doit être représentative de l'ensemble des configurations géométriques accessibles de la langue. Les valeurs des paramètres de contrôle doivent donc répondre à deux exigences : ils doivent permettre d'accéder à toutes les géométries de langue possibles, y compris les plus extrêmes et y compris les consonnes, tout en assurant une densité suffisante dans leur représentation, ceci afin de couvrir le mieux possible l'espace articulatoire. Pour cela nous avons fait 8800 simulations. Chaque simulation est générée à partir de la position de la langue au repos. La variable temporelle des commandes musculaires spécifie la durée de la transition (0.05ms) entre la position de repos et la position cible ainsi que la durée de tenue (0.15ms) de cette position cible.

4.2.2 Génération des patrons formantiques (F_1 , F_2 , F_3)

Nous rappelons que nous avons pris l'hypothèse selon laquelle la tâche de parole serait essentiellement spécifiée dans un domaine acoustico-perceptif et non articulatoire ou moteur (voir chapitre 3). Pour caractériser l'espace acoustico-perceptif de la tâche, nous avons choisi de décrire le spectre du signal de parole par les 3 premiers formants, F_1 , F_2 et F_3 . L'étape suivante de la génération de nos bases de données a donc consisté à calculer ces formants à partir des formes linguales générées par la méthode de Monte-Carlo décrite ci-dessus. Pour cela il est nécessaire de passer d'une description géométrique bi-dimensionnelle dans le plan sagittal de la tête à une description volumique du conduit vocal (appelée traditionnellement fonction d'aire) à partir de laquelle il sera possible de calculer les modes de résonances. Pour cela nous avons utilisé une version améliorée du modèle de passage de la coupe sagittal à la fonction d'aire proposé par Perrier (Perrier *et al.*, 1992).

Ce modèle caractérise de manière discrète la coupe sagittale du conduit vocal en calculant les intersections de la forme de ce conduit dans le plan sagittal avec une grille dont les lignes sont essentiellement perpendiculaires à la ligne principale de propagation de l'onde

acoustique depuis la glotte jusqu'aux lèvres (cf. figure 4.2). Le passage 2D – 3D repose sur l'hypothèse que la section transversale du conduit vocal, A , peut s'exprimer comme une fonction non-linéaire de la distance sagittale (distance séparant les contours externe et interne de la langue mesurée selon cette grille), d , selon l'expression : $A = c.d^{1.5}$, où c est un coefficient qui varie selon la position de la ligne glotte-lèvres dans le conduit vocal. Les valeurs du coefficient c proposées par Perrier *et al.* (1992) ont été modifiées par Maraoui (Maraoui, 2004) de façon à avoir des patrons formantique (F_1 , F_2 , F_3) pour les 3 voyelles extrêmes /i, a, u/ qui soient aussi proches que possible des données classiquement publiées dans la littérature (Calliope, 1989).

Les formants sont calculés à partir de la fonction d'aire en utilisant un analogue harmonique du conduit vocal modélisant ce conduit par une suite de tubes de sections variables, dont la fonction de transfert est calculée par analogie acoustico- électrique (Badin and Fant, 1984, voir aussi pour des exposés plus généraux sur la théorie de la modélisation acoustique de la parole, Fant, 1960 et Stevens, 1998).

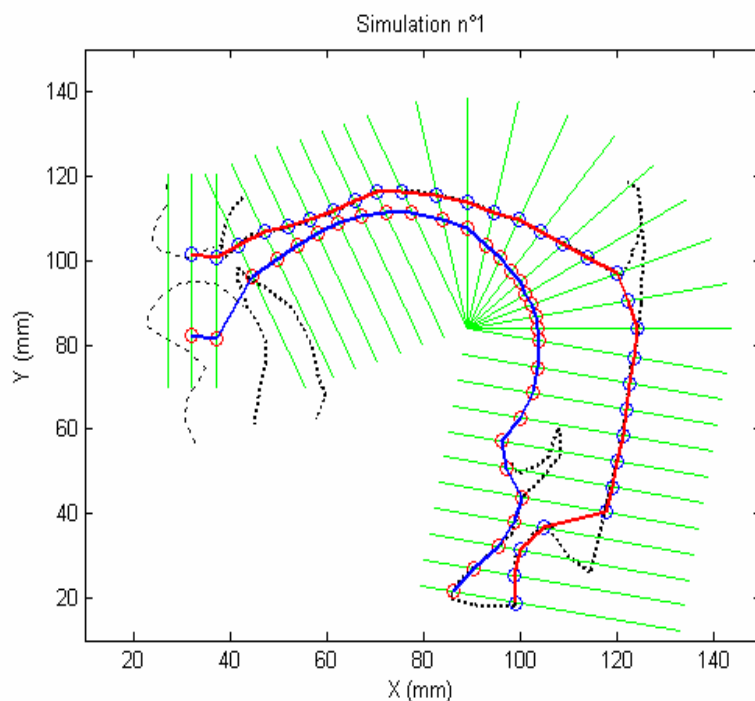


Figure 4.2 Contour du conduit vocal et grille de mesure de la coupe sagittale
Lèvres à gauche--glotte en bas à droite
Contour supérieur : parois du conduit vocal. Contour inférieur : langue

Le modèle articulatoire que nous avons utilisé est un modèle de la langue uniquement. Il ne possède en particulier aucune commande pour l'aire aux lèvres ni pour la mandibule. Afin

de pouvoir malgré cela générer des sons produits avec des lèvres ouvertes et des lèvres fermées, nous avons calculé pour chaque coupe sagittale du conduit vocal les patrons formantiques pour deux sections aux lèvres. Les lèvres ont été ainsi modélisées par un tube qui peut être soit long et étroit (50mm×20mm) pour des lèvres dites « arrondies » (/u/ par exemple), soit court et large (300mm×10mm) pour des lèvres dites « ouvertes » qui permet de générer toutes les voyelles du type (/i/, /e/, /ε/, /a/, /œ/). La figure 4.3 montre la fonction d'aire obtenue pour des lèvres ouvertes et pour le contour du conduit vocal de la figure 4.2, ainsi que les formants associés. Cette voyelle correspond à un /i/.

La figure 4.4 montre, dans le plan sagittal, les distributions des nœuds situés sur le contour externe du modèle de langue pour les 8800 simulations considérées. On peut constater que l'espace articulatoire est bien couvert avec des articulations consonantiques dans la partie vélaire et dans la partie palato-alvéolaire, sans qu'aucune forme linguale ne soit favorisée de manière évidente par rapport à une autre.

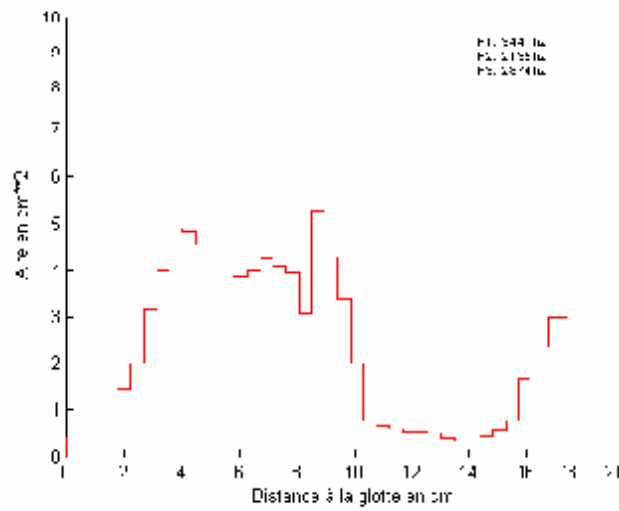


Figure 4.3 Fonction d'aire correspondant au contour de la figure 4.2, pour des lèvres ouvertes. Les valeurs correspondantes pour les 3 premiers formants sont présentées.

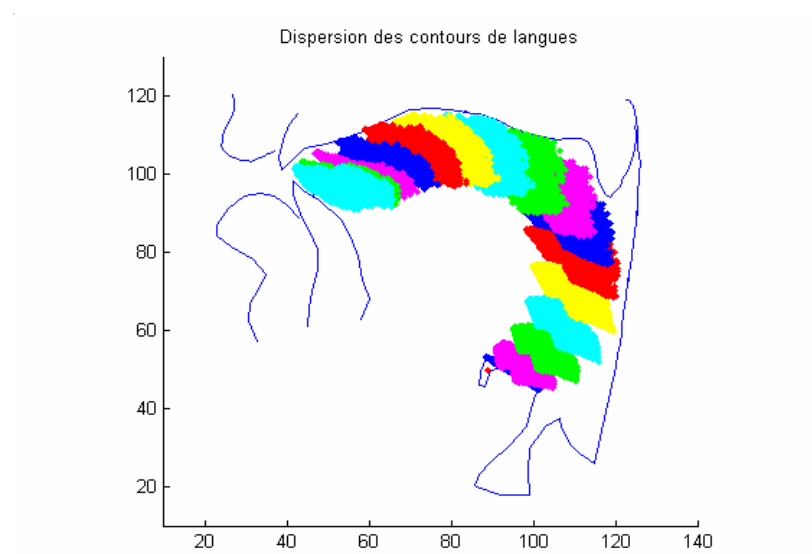
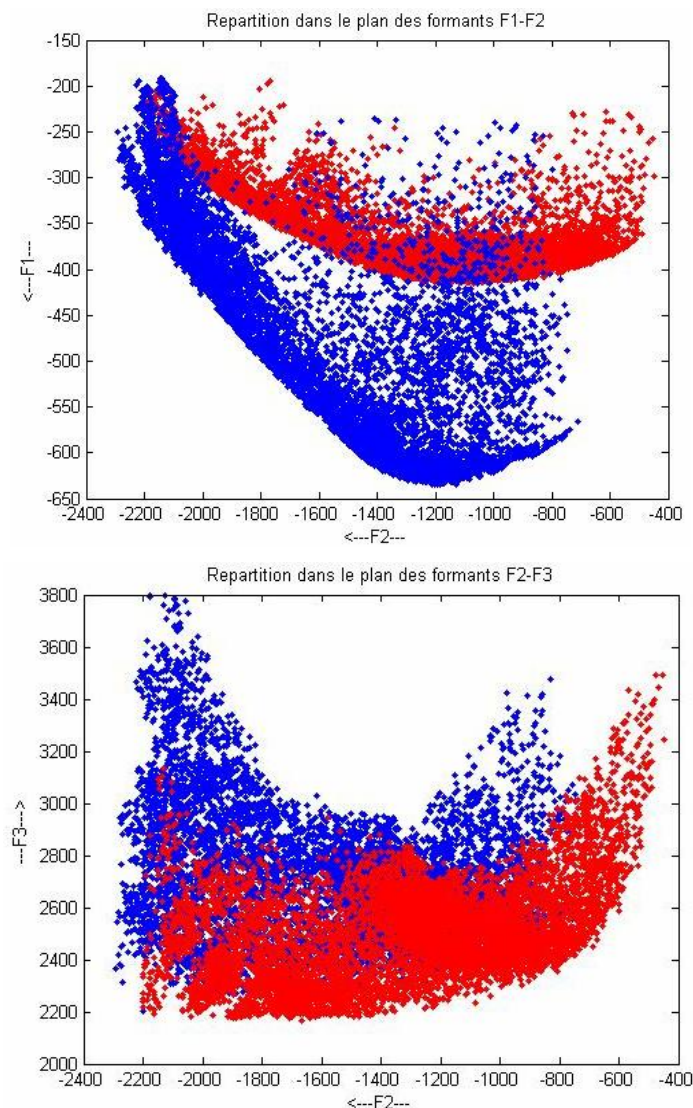


Figure 4.4 Répartition des 8800 simulations dans le plan sagittal. Chaque nuage de points correspond aux positions atteintes par chaque point du contour supérieur de la langue dans le modèle.

La figure 4.5 représente la distribution des simulations dans les plans (F_2, F_1) et (F_2, F_3) . On observe bien deux distributions de points différentes dans chaque plan : chacune d'entre elles correspond à une des deux formes de lèvres utilisées.



**Figure 4.5 Répartition des 8800 simulations dans les plans des formants (F_2, F_1) en haut et (F_2, F_3) en bas.
Rouge : simulations pour des lèvres fermées Bleu: simulations pour des lèvres ouvertes**

Contrairement à la distribution des données dans le plan moteur puis dans le plan articulatoire (cf. Figure 4.4), la distribution des patrons formantiques n'est plus uniforme avec, pour des lèvres ouvertes une représentation beaucoup plus dense le long d'un axe /i, e, ε, a/ dans le plan (F_2, F_1) et pour des lèvres fermées, une densité plus importante le long de la direction /y/-/ø/-/u/ dans ce même plan. Ceci est dû à la non-linéarité entre l'espace articulatoire et l'espace acoustique. Il laisse présager des jeux de compensation et une variété

des stratégies articulatoires beaucoup plus importants pour ces voyelles que pour les voyelles plus centrales. Ce phénomène a déjà été observé par Maeda (1979) avec un modèle articulatoire géométrique qui devrait ainsi générer un espace beaucoup plus dense selon l'axe /i, e, ε, a/ que dans le reste du triangle des voyelles.

4.3 Modèle interne statique

4.3.1 Introduction d'apprentissage du modèle interne statique

A partir de ces données associant commandes motrices et formants F_1 , F_2 , F_3 , que nous avons générées avec le modèle biomécanique, nous avons élaboré un modèle fonctionnel statique des relations entre l'espace des commandes motrices et l'espace acoustico-perceptif : un modèle interne statique. Dans la lignée des hypothèses de Jordan (1990), Kawato *et al.* (1990), Hirayama *et al.* (1992), et Jordan et Rumelhart (1992), ce modèle correspond à une représentation que le système nerveux central aurait de ces relations et qu'il exploiterait pour élaborer les séquences motrices permettant la production d'une séquence de parole (voir chapitre 3). Dans la mesure où nous n'avons que deux formes prototypiques des lèvres, il n'est pas statistiquement possible de prendre en compte le paramètre de commande labiale dans cet apprentissage. Par ailleurs, nous n'avons pas de modèle physique réaliste des lèvres qui serait l'équivalent du modèle de la langue. Nous n'avons donc pas envisagé de prendre en compte de manière impérative les mouvements labiaux. Nous avons entraîné le modèle avec les simulations obtenues avec les lèvres ouvertes (Figure 4.6) et seuls les résultats associés à cette configuration labiale seront présentés dans cette étude.

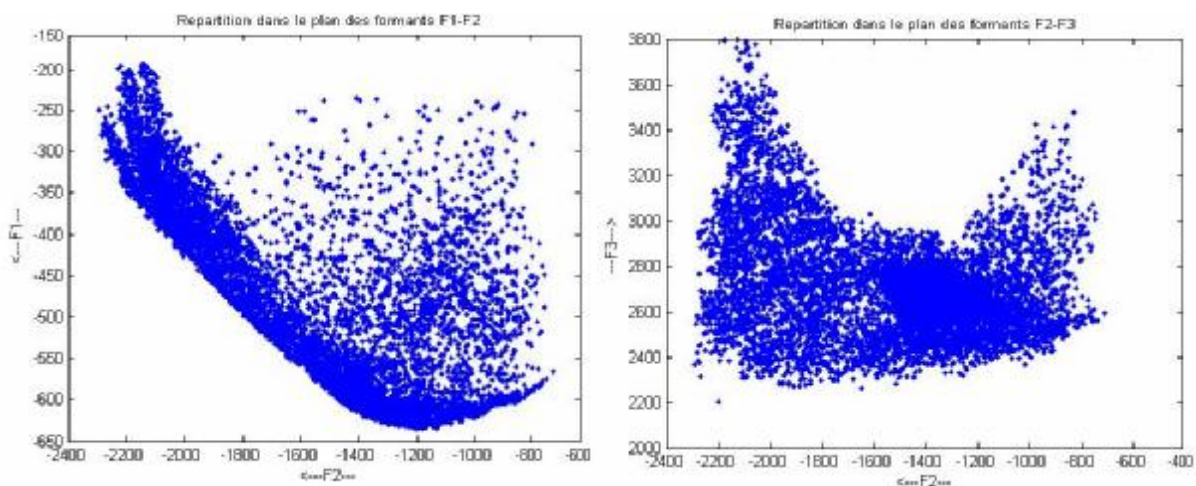


Figure 4.6 Répartition des simulations avec les lèvres ouvertes dans les plans des formants (F_2 , F_1) et (F_2 , F_3).

La fonction qui caractérise les relations entre l'espace de commandes motrices et l'espace acoustico-perceptif n'est pas bijective, puisque d'une part, du fait des synergies et antagonismes musculaires, une même forme de la langue peut en principe être associée à différents jeux de commandes motrices et que, d'autre part, selon la théorie acoustique de la production de la parole, une même combinaison (F_1 , F_2 , F_3) est potentiellement associée à différentes formes de la langue. De plus, cette relation est hautement non linéaire tant à cause du modèle de génération de force de Feldman (1986) qu'à cause des non-linéarités de la relation entre la géométrie du conduit vocal et les valeurs formantiques (Stevens, 1989). Pour modéliser de manière fonctionnelle des relations de ce type, des fonctions radiales de bases semblent être très efficaces (Poggio et Girosi, 1989). Notre travail sur le modèle interne est la suite d'un premier travail de Marret et Perrier (Marret, 2002 ; Perrier *et al.*, 2004). Ils ont montré une démarche préliminaire pour l'apprentissage de formes purement vocaliques. Nous avons employé les programmes développés par Marret (2002) pour apprendre sur notre base de simulation les paramètres des Fonctions Radiales de Bases permettant la meilleure description fonctionnelle des relations entre commandes motrices et formants.

4.3.2 Les fonctions radiales de bases

Les propriétés mathématiques caractérisant les fonctions radiales de base ont été décrites en détail par Poggio et Girosi (1989). Il existe un grand nombre de fonctions différentes qui satisfont ces propriétés. Pour notre part, conformément à ce qui est classiquement fait dans la littérature, nous utiliserons des noyaux de type gaussien. La valeur de la fonction dépend alors de la distance entre le vecteur d'entrée P et le vecteur W caractérisant les centres de ce noyau gaussien selon la relation :

$$f(P) = e^{-\frac{\|P-W\|^2}{\sigma^2}}$$

où σ est le vecteur de la variance du noyau gaussien.

Nous avons pris ici le vecteur d'entrée P à 6 dimensions qui correspondent aux 6 commandes des muscles utilisés dans notre base de simulations.

Concrètement pour l'implémentation du modèle interne, la fonction Matlab appelée *radbas* est utilisée. Le schéma équivalent dans un espace à R dimensions et l'expression sont présentés en figure 4.7.

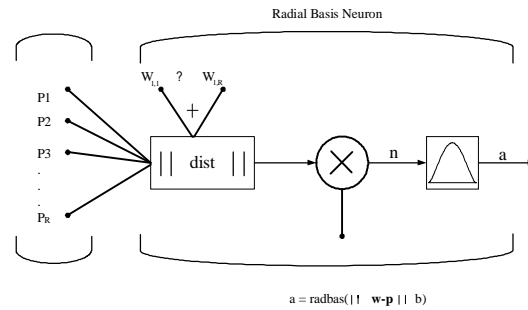


Figure 4.7 Schéma de principe d'une fonction radiale de base à noyau gaussien.

La fonction *radbas* a un maximum de 1 et son écart-type est aussi normalisé à 1. Pour jouer sur la valeur effective de l'écart-type de la fonction radiale de base que nous utiliserons, il faut modifier la valeur du paramètre de biais b , selon la formule $\sigma=1/b$.

Le schéma du réseau de neurones proposé par Marret (2002) est présenté en figure 4.8. Il s'agit d'un réseau à deux couches. La première, dite couche cachée, calcule pour le noyau gaussien d'indice j la distance entre le vecteur d'entrée (appelé ici X) et le vecteur 1W_j centre de ce noyau, puis calcule la sortie de la fonction radiale de base (notée F) correspondante. La seconde couche, dite couche de sortie, est simple : elle se contente de calculer la somme pondérée (poids 2W_j) des sorties des noyaux gaussiens de la première couche.

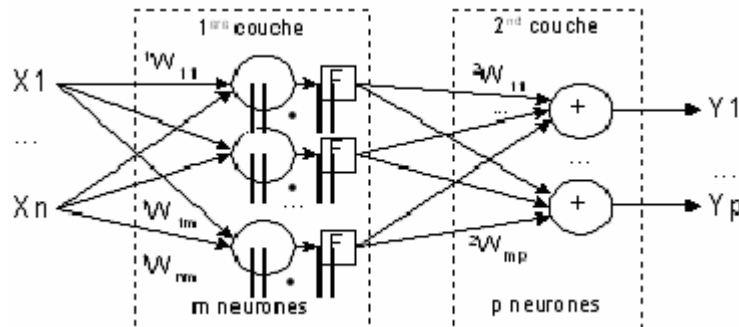


Figure 4.8 Réseau de neurones à deux couches utilisé pour apprendre le modèle interne statique (Marret, 2002).

La fonction synthétisée par le réseau est définie ainsi (Marret, 2002) :

$$f^{\#}(X) = \sum_{j=1}^J {}^2W_j g_{j_i}(x) \text{ où } J \text{ est le nombre de gaussiennes, et où } \{g_j\}_{j=1..J} \text{ est l'ensemble}$$

des sorties des noyaux gaussiens du réseau.

4.3.3 Principe de l'apprentissage

Ce type de réseau présente de bonnes caractéristiques autant au niveau de la qualité de l'approximation effectuée que de la généralisation obtenue ou de la régularité de la solution. L'apprentissage consiste à trouver les valeurs optimales pour (1) le nombre m des noyaux gaussiens, (2) les positions 1W_j ($j=1\dots m$) de leurs centres, (3) les valeurs de leur biais b_j ($j=1 \dots m$), et (4) les poids $^2W_{jp}$ ($j=1\dots m$) ($k= 1..p$) combinant linéairement les sorties des noyaux gaussiens d'indice j pour trouver la sortie d'indice p . Cette optimisation est très délicate à mettre en oeuvre, principalement parce que la qualité de l'apprentissage est fortement conditionnée par la position initiale des centres des gaussiennes et par la valeur de leur écart type. Les algorithmes qui adaptent l'écart type de chaque gaussienne sont complexes et coûteux en temps de calcul, et ils ne sont pas pris en charge par Matlab. Marret n'a donc pas utilisé un tel algorithme et il s'est contenté de tester différentes valeurs du paramètre b maintenu fixe pendant chaque session d'apprentissage. La condition à respecter pour ce paramètre b est de choisir une valeur telle que l'écart type correspondant soit suffisamment grand afin que les régions actives des noyaux gaussiens se chevauchent pour garantir un caractère lisse à la fonction ainsi générée, tout en ayant un degré de granularité suffisant pour assurer une approximation correcte des relations entrées-sorties

L'objectif d'établir un modèle interne pour nous est de décrire une relation généralisée entre les commandes motrices et les trois premiers formants de nos données de base. Nous mettons ainsi en relation le vecteur d'entrée à 6 dimensions qui correspondent aux 6 commandes musculaires, et le vecteur de sortie à trois dimensions qui correspondent aux trois premiers formants. Pour apprendre et généraliser les informations contenues dans cette base, nous avons subdivisé l'ensemble des 8800 données générées en deux sous-ensembles appelés respectivement base d'apprentissage et base de test.

La fonction d'erreur à minimiser dans l'algorithme est l'erreur quadratique sur les formants (F_1, F_2, F_3). L'algorithme d'apprentissage que Marret a utilisé est fourni dans la boîte à outils Neural Networks de Matlab. L'algorithme s'applique à la base d'apprentissage et consiste en les étapes suivantes : il part d'un nombre de noyaux gaussiens arbitraire; puis l'algorithme choisit les valeurs optimales des centres de ces noyaux, qu'il positionne sur une des données de la base d'apprentissage, ainsi que des poids optimaux de la couche de sortie ; les erreurs quadratiques ainsi obtenues sur la base d'apprentissage et sur la base de test sont calculées; ensuite l'algorithme ajoute un noyau gaussien et recommence le processus complet d'optimisation.... Pour déterminer le moment où il faut arrêter ce processus, Marret a proposé

d'observer simultanément l'évolution de l'erreur sur la base d'apprentissage et sur la base de test. Classiquement si l'algorithme est efficace, l'erreur sur la base d'apprentissage diminue systématiquement. C'est le cas aussi de l'erreur sur la base de test, jusqu'à un certain point où elle alors se met à augmenter de nouveau. Ce point est le point crucial de l'apprentissage. Il correspond au moment où la capacité de généralisation du réseau est maximale. C'est à ce point donc que nous arrêterons le processus d'apprentissage.

Nous avons enfin choisi 5000 simulations de nos données de base pour l'apprentissage et 3800 pour le test. La figure 4.9 montre l'évolution des erreurs quadratiques d'apprentissage et de test. 400 noyaux gaussiens ont été retenus ici : au-delà l'erreur de test reste essentiellement constante avant de remonter autour de 600 noyaux. Compte tenu du faible gain de précision obtenu par le passage de 400 à 600 noyaux, nous avons retenu ce premier nombre.

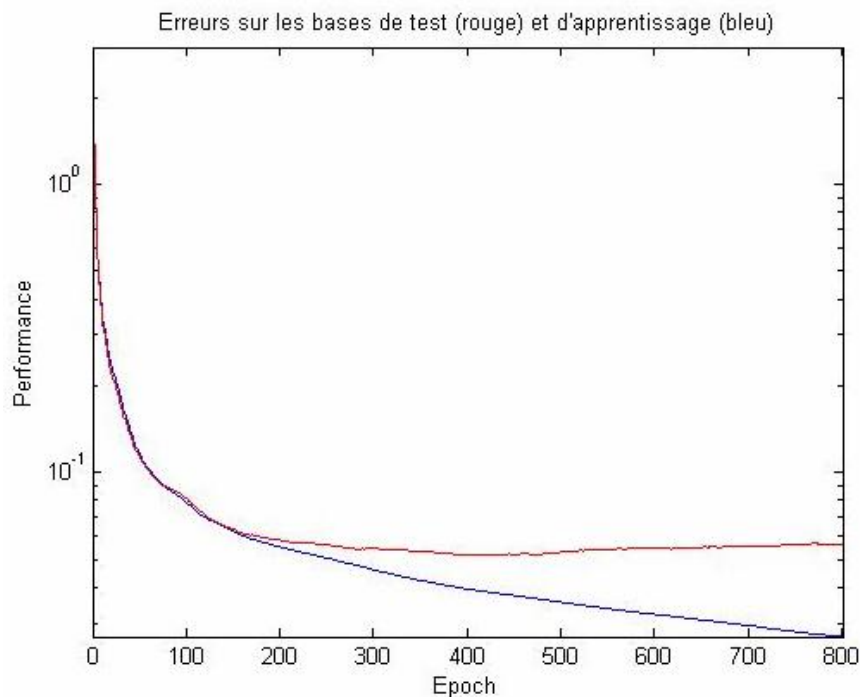


Figure. 4.9 Evolution des erreurs sur les bases d'apprentissage (courbe inférieure) et de test (courbe supérieur)

La figure 4.10 présente les résultats. Chaque tracé représente en abscisse la valeur de F_1 (en haut à gauche), F_2 (en haut à droite) et F_3 (en bas) à la fois pour la base de test et pour la base d'apprentissage de manière cumulée, et en ordonnée la valeur de sortie correspondante du réseau. La diagonale correspond à des points parfaitement appris, le nuage de point est donc réparti autour de cette diagonale. Les deux droites tracées correspondent à une marge

d'erreur de $\pm 10\%$ sur la valeur théorique. Le pourcentage de la base de test restitué avec une erreur inférieure à 10.0% est pour F_1 : 99.7% , pour F_2 : 99.9% , pour F_3 : 99.7% . Avec notre critère d'erreur, nous pouvons donc considérer que notre réseau de neurones a bien décrit les données générées par le modèle et qu'il a su généraliser.

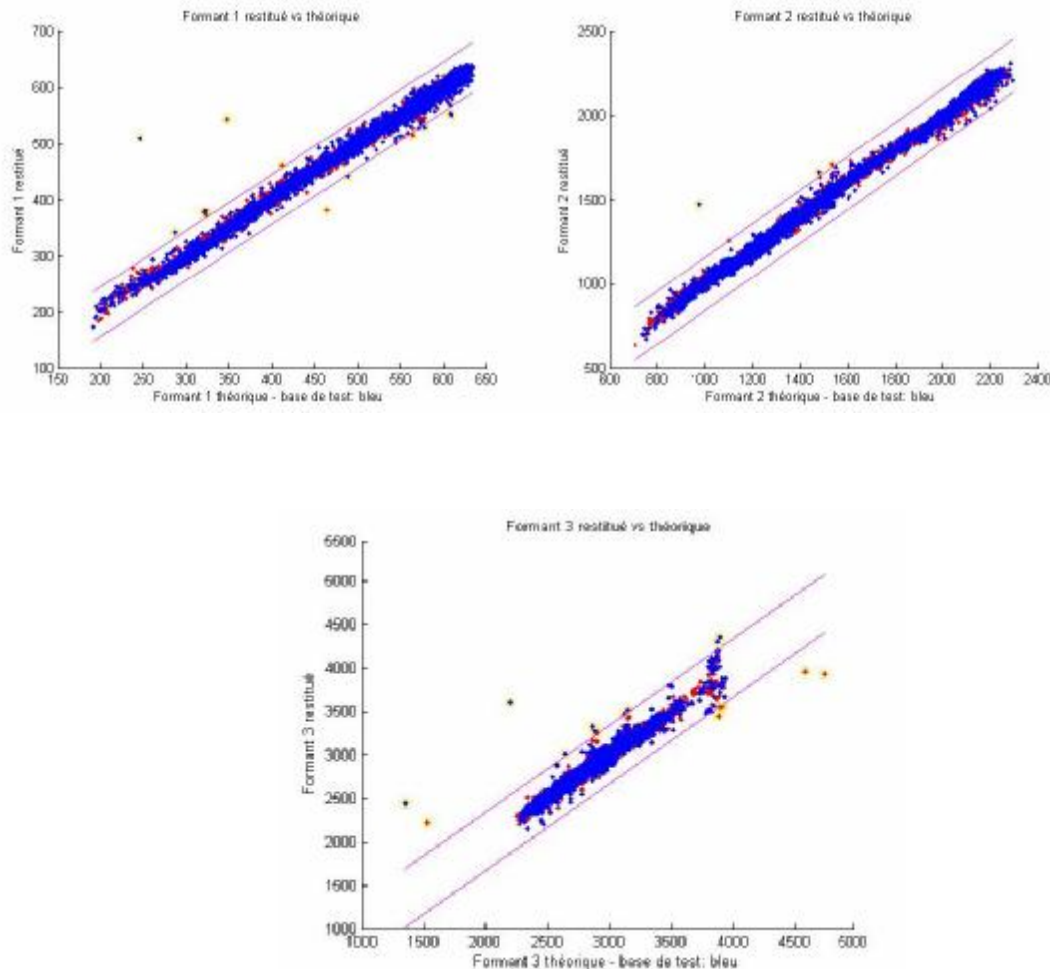


Figure4.10 Qualité de l'apprentissage des trois premiers formants

4.4 Inversion : des commandes phonémiques aux commandes motrices

Le modèle interne est établi grâce aux fonctions radiales de bases. Ce modèle sera ensuite utilisé afin d'exploiter les synergies et les antagonismes musculaires lors de la planification gestuelle d'une séquence de parole grâce à la combinaison d'une planification gestuelle optimale et d'une contrainte perceptive. La phase de planification de la séquence consiste à inverser les commandes motrices associées à la génération des caractéristiques spectrales cibles souhaitées pour le signal de parole. Cette inversion n'est pas unique car : (1) les caractéristiques spectrales d'un même phonème ne sont pas uniques et (2) les relations entre commandes motrices et caractéristiques spectrales du signal acoustique ne sont pas

biunivoques. Il n'y a donc pas de solution analytique au problème et la recherche d'une solution passe donc par une étape de régularisation. Nous avons choisi pour cela l'optimisation d'un critère sur l'ensemble de la séquence. Une première proposition de critère (Maret, 2002) a été testée, qui intègre à la fois des contraintes orientées vers le locuteur (minimisation d'un « effort ») et des contraintes orientées vers l'auditeur (maximisation de l'efficacité perceptive du son à la cible).

4.4.1 Corrélats acoustico-perceptifs des phonèmes

Désormais pourvu d'une représentation interne assez précise des relations directes entre commandes motrices et caractéristiques formantiques, le modèle de contrôle de la production de la parole doit être capable de trouver les commandes motrices sous-jacentes à la production de n'importe quelle séquence de parole. Pour cela nous allons faire une hypothèse consistant à dire que tous les sons, y compris les consonnes, sont spécifiés par des caractéristiques particulières dans l'espace (F_1 , F_2 , F_3). Pour cela nous proposons que soit associée à chaque phonème une zone spécifique de l'espace (F_1 , F_2 , F_3), que nous appellerons *région cible*, et qui est pleinement définie sous la forme de deux ellipses de dispersion, l'une dans le plan (F_2 , F_1) et l'autre dans le plan (F_2 , F_3), dont on spécifie les centres F_{c1} , F_{c2} , F_{c3} et les écart-types σ_{F1} , σ_{F2} , σ_{F3} . Cette démarche est dans la lignée des propositions de Keating (1988) dans son modèle de fenêtre et reprises par Guenther (1995, 1998) dans son modèle de DIVA. Pour définir ces paramètres pour les voyelles, nous nous sommes largement inspirés de données publiées sur le français dans la littérature (Calliope, 1989).

Pour les consonnes, nous avons opéré de la manière suivante. Nous avons calculé les formants à l'aide du modèle harmonique du conduit vocal pour toutes les formes du conduit vocal que nous avons générées avec le modèle, y compris pour celles qui présentent un fort rétrécissement, voire une occlusion. Ainsi il nous a été possible de voir quelles étaient les zones de l'espace (F_1 , F_2 , F_3) qui étaient associées par ce modèle harmonique à des occlusions dans la partie vélaire (consonnes plosives /k, g/), ou dans la partie alvéolaire (consonnes plosives /t, d/), ou à de forts rétrécissements dans ces mêmes régions (consonnes fricatives /s, z, ʃ, ʒ, /). Ce sont ces régions, formellement décrites sous forme d'ellipses de dispersion, que nous avons prises en compte pour caractériser les consonnes dans le domaine spectral. La figure 4.11 montre la zone sélectionnée pour la consonne vélaire /k/ ou /g/ dans le plan ayant en abscisse la position le long d'une ligne glotte-lèvres et en ordonnées l'aire du conduit vocal à la constriction.

Pour définir cette cible de la consonne plosive dans le plan formantique, nous avons constaté que la caractérisation perceptive des consonnes classiquement proposée dans la littérature repose uniquement sur la fréquence du bruit et/ou sur les directions des transitions formantiques et non pas, comme nous le proposons, sur un point fixe dans l'espace formantique pendant la tenue consonantique. Cependant, nous pouvons confirmer, comme l'a fait Galvan (1997) que ce point fixe pourrait être théoriquement relié au *locus*, ce point virtuel de l'espace formantique d'où seraient issues les transitions formantiques de la voyelle consécutive à la consonne (Delattre *et al.*, 1955 ; Sussman *et al.*, 1997). De plus, même si les spectres calculés ne sont pas réalistes, ils peuvent être utilisés dans le cadre limité du modèle acoustique que nous utilisons pour caractériser articulatoirement les formes consonantiques.

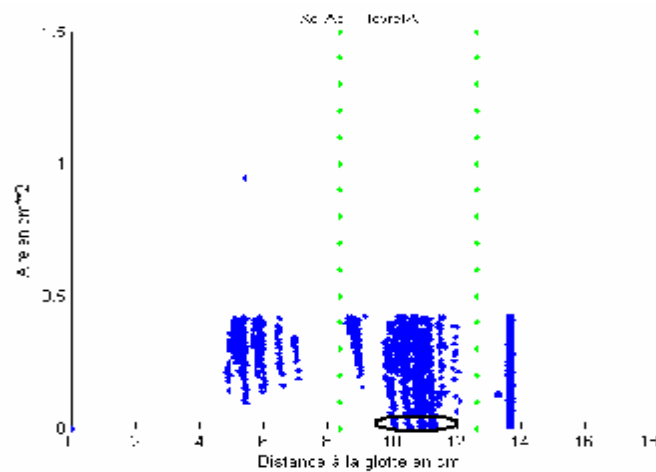


Figure 4.11 Zone retenue (ellipse en trait gras) pour la consonne plosive vélaire /k/, /g/ dans le plan (Distance de la constriction à la glotte – Aire à la constriction).

Les valeurs que nous avons choisies pour caractériser les voyelles orales et les consonnes, dans le cas de lèvres ouvertes et donc non arrondies, sont données dans le tableau 1.

	/i/	/e/	/ɛ/	/a/	/œ/	/ɔ/	/k/
F _{c1}	288	375	508	614	498	550	276
F _{c2}	2178	2097	1689	1214	1508	1019	1190
F _{c3}	3005	2413	2390	2640	2566	2499	2614
σ _{F1}	10	30	40	15	40	30	15
σ _{F2}	110	70	80	130	120	60	350
σ _{F3}	140	90	100	130	100	100	200

Tableau 1 : Centres et écart-types des ellipses de dispersion caractérisant les voyelles orales non arrondies /i, e, ɛ, a, œ, ɔ/ et la consonne vélaire /k/.

4.4.2 Inversion du modèle interne par optimisation d'un critère

Pour trouver la séquence des commandes motrices cibles sous-jacentes à la production d'une séquence de parole donnée, le principe est de minimiser une fonction du coût qui a été élaborée afin de prendre en compte deux critères (Marret, 2002 ; Perrier *et al.*, 2004).

- Le premier critère dit "*orienté vers le locuteur.*"
- Le deuxième critère dit "*orienté vers l'auditeur.*"

Le premier critère correspond à une prise en compte de l'effort du locuteur. Pour cela, la minimisation du chemin parcouru pendant la séquence par les commandes cibles dans l'espace des commandes musculaires a été choisie. L'idée est de générer le minimum de déplacement articulaire.

Le second critère a pour objectif de garantir que l'effet perceptif produit est correct. Pour cela, il faut impérativement que le signal produit ait ses caractéristiques formantiques à l'intérieur de la région cible définie pour le phonème considéré dans l'espace formantique (F_1 , F_2 , F_3). Un critère perceptif est donc défini. Il prend une valeur très faible tant que (F_1 , F_2 , F_3) reste dans la région cible recherchée, et devient très grand dès que les caractéristiques acoustiques sortent de cette région.

L'objectif de l'inversion est de minimiser le premier critère dans l'espace de commandes motrices en prenant en compte la contrainte du deuxième critère dans l'espace perceptif. Pour atteindre cette minimisation, l'algorithme du gradient a été utilisé. L'utilisation du gradient a été retenue car il s'agit d'une approche classique dans la problématique du contrôle moteur (Jordan et Rumelhart, 1992 ; Kawato *et al.*, 1990). La minimisation de cette fonction de coût permet de rechercher une solution optimale, pour laquelle les cibles phonétiques choisies correspondent perceptivement bien aux sons que l'on veut produire, tout en minimisant la variation globale des commandes musculaires cibles pendant toute la séquence.

4.4.3 Exemples d'inversion

Nous avons, dans un premier temps, évalué ce modèle global en testant sa capacité à trouver les commandes motrices adaptées à la production de séquences (Voyelle-Consonne-Voyelle). Dans le modèle biomécanique, il est nécessaire, pour des raisons computationnelles, que, quelle que soit la séquence de parole synthétisée, la position initiale de la langue soit systématiquement dans sa position de repos, qui est unique et ne peut donc pas être partie

intégrante de la planification. Par exemple, pour générer la séquence /a-k-e/ nous avons synthétisé la séquence /repos-a-k-e/.

Les figures 4.12 et 4.13 présentent les résultats de l'inversion pour cette séquence /repos-a-k-e/, respectivement dans les plans (F2, F1) et (F2, F3). On observe dans chaque cas 3 ellipses, celles au sein desquelles vont se situer les cibles des 3 sons élémentaires autres que la position de repos /a/, /k/, /e/. Les lignes droites qui relient ces ellipses, connectent entre elles les positions cibles associées, selon le modèle interne, aux commandes motrices sélectionnées par le processus de planification optimale. Au début de l'algorithme d'optimisation, les commandes motrices correspondant au centre de chacune des ellipses sont choisies pour représenter le phonème désiré. Ces premières commandes motrices vont générer un premier point dans l'espace formantique. Ce point n'est plus le point au centre de chaque ellipse dans le plan formantique à cause du modèle interne. Grâce à l'itération d'algorithme, le chemin optimal entre commandes cibles est obtenu petit à petit. La figure 4.14 présente les 6 commandes musculaires des muscles Génio-glosse Postérieur (GGp), Génio-glosse antérieur (Gga), Styloglosse (Sty), Hyoglosse (Hyo), Verticalis (Ver) et Longitudinalis inférieur (IL) qui varient pendant les itérations jusqu'à la convergence. Nous sommes intéressés par la dernière itération qui correspond donc aux commandes musculaires optimales pour générer notre séquence /repos-a-k-e/.

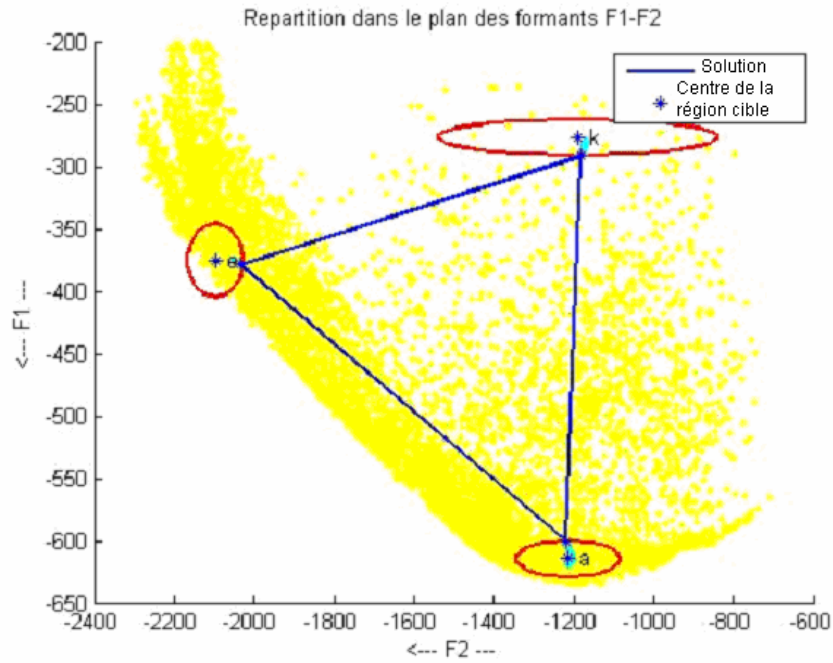


Figure 4.12 Plan des formants F_1 et F_2

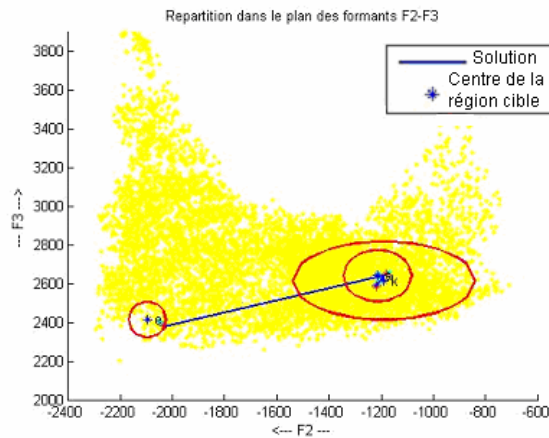


Figure 4.13 Plan des formants F_2 et F_3

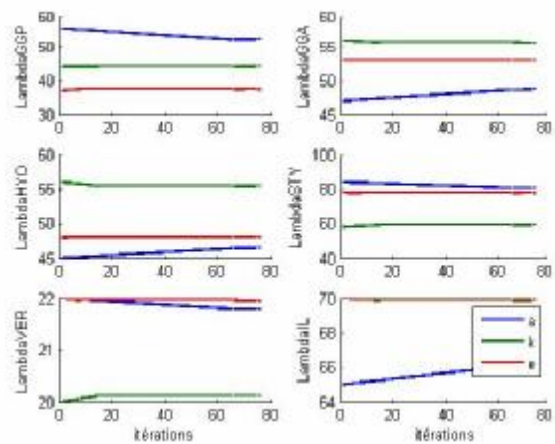


Figure 4.14 commandes musculaires

Les figures 4.15, 4.16 et 4.17 présentent les mêmes résultats pour la séquence /repos-a-k-ə/. Si nous comparons les deux figures 4.12 et 4.15, nous observons un effet de la planification : le point choisi dans l'ellipse du /a/ et dans l'ellipse du /k/ correspond à un F_2 plus faible pour /repos-a-k-ə/ que pour /repos-a-k-e/. Nous observons donc un effet d'anticipation dû à la planification : nous voyons dans la commande cible de /a/ et de /k/ apparaître le fait que la voyelle suivante sera plus postérieure (F_2 plus faible).

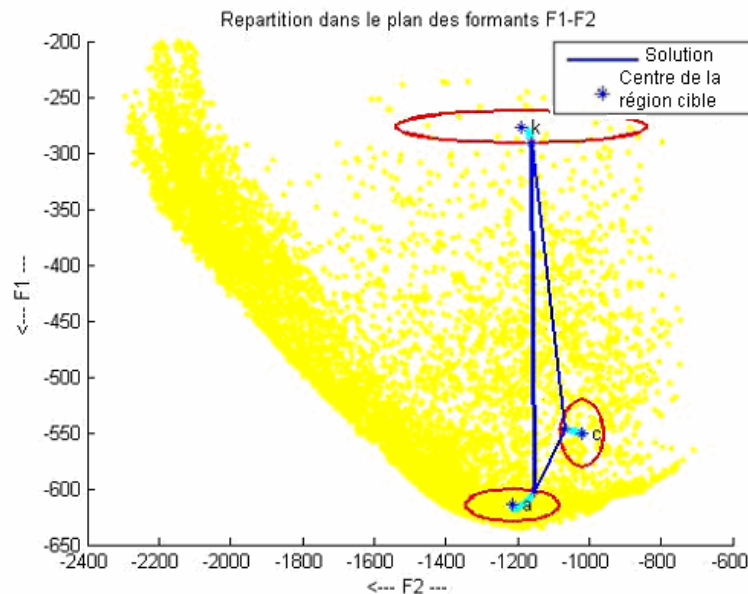


Figure 4.15 Plan des formants F_1 et F_2

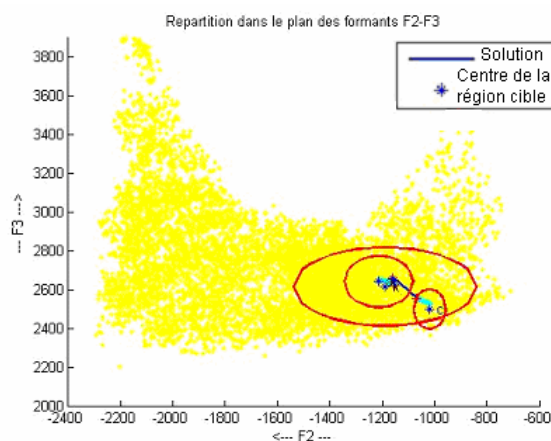


Figure 4.16 Plan des formants F_2 et F_3

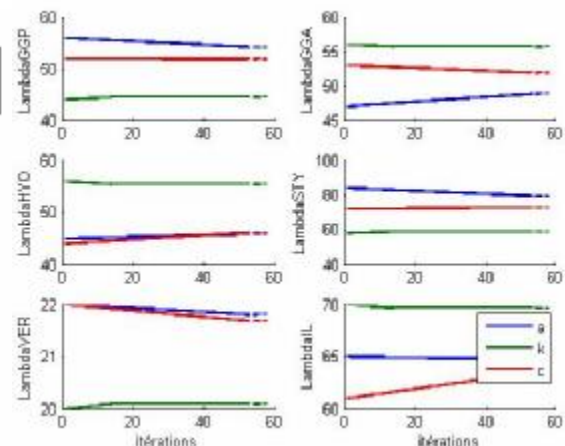


Figure 4.17 commandes musculaires

Chapitre 5 : Méthodologie Expérimentale

Introduction

Comme nous l'avons dit dans le chapitre de problématique (chapitre 4), nous allons dans le but d'étudier l'influence linguistique sur la planification de la parole, analyser la coarticulation d'anticipation sur les données articulatoires en français et en chinois. Ce chapitre décrit la méthode utilisée pour recueillir et traiter les données articulatoires. Après une description des données recueillies et une description du corpus utilisé, nous présenterons la méthode développée pour repérer les instants les plus caractéristiques de l'articulation des sons successifs de la séquence. A la fin de ce chapitre, nous présenterons également la modification du contour du palais pour chaque locuteur.

5.1 Recueil et description des données

5.1.1 Recueil des données

Les données articulatoires de cette étude ont été enregistrées grâce au système EMMA (2D). (AG100- Carstens Electronics). Ce système de mesure par induction électromagnétique consiste d'une part en un casque sur lequel sont fixées trois bobines électriques et en des capteurs, petites bobines de quelques spires, qui sont collées sur les articulateurs dont on veut analyser le mouvement. Les trois bobines du casque, placées aux angles d'un triangle équilatéral, émettent des champs magnétiques sinusoïdaux de fréquences différentes. L'intensité du courant ainsi induit dans les capteurs est inversement proportionnelle au cube de la distance entre le capteur et la bobine émettrice et informe donc en permanence de la position du capteur.

Le système EMMA bi-dimensionnel exige un placement de l'ensemble des capteurs dans le plan sagittal de la tête. Quatre capteurs ont été ainsi placés sur la langue, un capteur sur chaque lèvre, un sur les incisives inférieures, un sur les incisives supérieures et enfin un sur l'arête du nez. Les deux derniers capteurs constituent nos points de référence. Les quatre capteurs placés sur la langue, répartis uniformément entre l'apex de la langue et une position située à environ 5 cm de l'apex, sont notés T_1 , T_2 , T_3 et T_4 . La fréquence d'échantillonnage est de 500 Hz pour les canaux articulatoires et de 20 kHz pour les données acoustiques. Pour le

sujet chinois JW, la fréquence d'échantillonnage est de 200 Hz.

5.1.2 Description du corpus

Le corpus est composé de 15 phrases. Il a été enregistré par trois sujets chinois et trois sujets français. Des séquences de type VCV sans signification ont été enregistrées à partir des phrases lues. Les voyelles utilisées étaient /a i u/ et les consonnes étaient la vélaire /k/ et l'alvéolaire /t/. Chaque mot cible a été inclus dans la phrase porteuse :

C'est VCV ça ? (En français)

这是 VCV 吗? (En chinois)

La transcription phonétique de la phrase en français est /sɛ/-VCV-/sa/ ? La transcription phonétique de la phrase en chinois est /tʂə tsɿ /-VCV-/ma/ ? Cette phrase en chinois signifie *c'est VCV ça ?* en français. Chaque phrase a été répétée 10 fois à l'exception d'un sujet (cf 5.1.3). Les données acoustiques et les données articulatoires ont été enregistrées simultanément.

Le corpus a été élaboré afin d'avoir un contexte phonétique comparable dans les deux langues, malgré des différences bien connues dans leur inventaire respectif de phonèmes ainsi que dans leurs structures linguistiques respectives (voir le chapitre 2). En particulier, en chinois nous avons évité les séquences qui pourraient être ambiguës en raison de la structure tonale. Toutes les séquences VCV en chinois sont produites avec le ton haut, c'est-à-dire le premier ton en chinois. Nous n'avons également pas considéré des séquences comme /V₁ki/ qui n'existent pas en chinois. Les séquences VCV ont été produites avec le même degré d'accentuation sur les deux syllabes pour éviter les modèles d'accentuation des différentes langues. Le corpus de séquences V₁CV₂ utilisées en français et en chinois est présenté dans le tableau 5.1.

	C /t/	C /k/
V ₁ /a/	/ata/ /ati/ /atu/	/aka/ /aku/
V ₁ /i/	/ita/ /iti/ /itu/	/ika/ /iku/
V ₁ /u/	/uta/ /uti/ /utu/	/uka/ /uku/

Tableau 5.1 Le corpus

5.1.3 Les sujets

Parmi les trois sujets français, il y a un sujet féminin (AV) et deux sujets masculins (PB et CV). AV viens de l'Est de la France, et PB et CV sont originaires du Sud de la France. Les trois sujets chinois sont masculins. SK vient du centre de la Chine (Shaanxi), WS vient du Sud-est de la Chine (Nanjing), et JW est originaire du centre de la Chine (Shaanxi) et a vécu au Japon pendant vingt ans.

Le sujet chinois JW a été celui de l'étude pilote à partir de laquelle le corpus définitif a été établi. C'est pour cela qu'il a seulement produit 4 répétitions des séquences VCV au lieu de 10. Les capteurs T₂ et T₃ situés sur la langue pour le sujet WS, le capteur T₃ du sujet AV, ainsi que le capteur T₁ du sujet SK et du sujet CV n'ont pas fonctionné pendant la totalité de l'expérience (rupture du câble à la jonction avec le capteur). Nous avons enfin enlevé tous ces capteurs dans notre analyse, nous ne présentons que les capteurs qui ont fonctionné.

5.2 L'étiquetage

Les données EMMA ont été filtrées par un filtre passe-bas à 18 hertz (-20dB et 20.1 Hz). Dans un premier temps, l'étiquetage des données a été effectué manuellement pour localiser des événements acoustiques. L'objectif est de repérer les positions acoustiques les plus caractéristiques des sons produits. Dans un deuxième temps, l'étiquetage manuel de la voyelle a été modifié automatiquement pour trouver la cible articulaire atteinte (le point le plus extrême des articulateurs).

5.2.1 Le logiciel pour l'étiquetage

Les données articulaires et acoustiques ont été traitées et étiquetées avec une interface utilisateur spécifiquement développée (Ma, 2004). Nous avons élaboré un logiciel pour l'étiquetage des données. Ce logiciel intègre un ensemble de fonctions de base du traitement du signal (Analyse de Fourier et Filtrage), ainsi que le calcul des dérivées première et seconde des signaux de mouvement. Ces signaux de mouvement ont été analysés après avoir été filtrés passe-bas (18 Hz). L'interface utilisateur a été élaborée de façon à pouvoir sélectionner les signaux à visualiser, à effectuer des zooms aux endroits appropriés et à positionner des étiquettes via des curseurs dont la position temporelle est mémorisée dans un fichier de résultats.

La figure 5.1 présente l'interface d'analyse des données. Il y a 5 signaux représentés et un spectrogramme: le premier signal correspond à la vitesse tangentielle d'un capteur, le deuxième représente la vitesse en Y de ce même capteur, le troisième et le quatrième sont les déplacements correspondant X et Y, le cinquième est le signal audio, dont le spectrogramme est représenté dans le cadre inférieur. Le capteur étudié peut être sélectionné par le menu (situé en haut : T₁, T₂, T₃, T₄). Nous allons expliquer comment faire l'étiquetage par la suite.

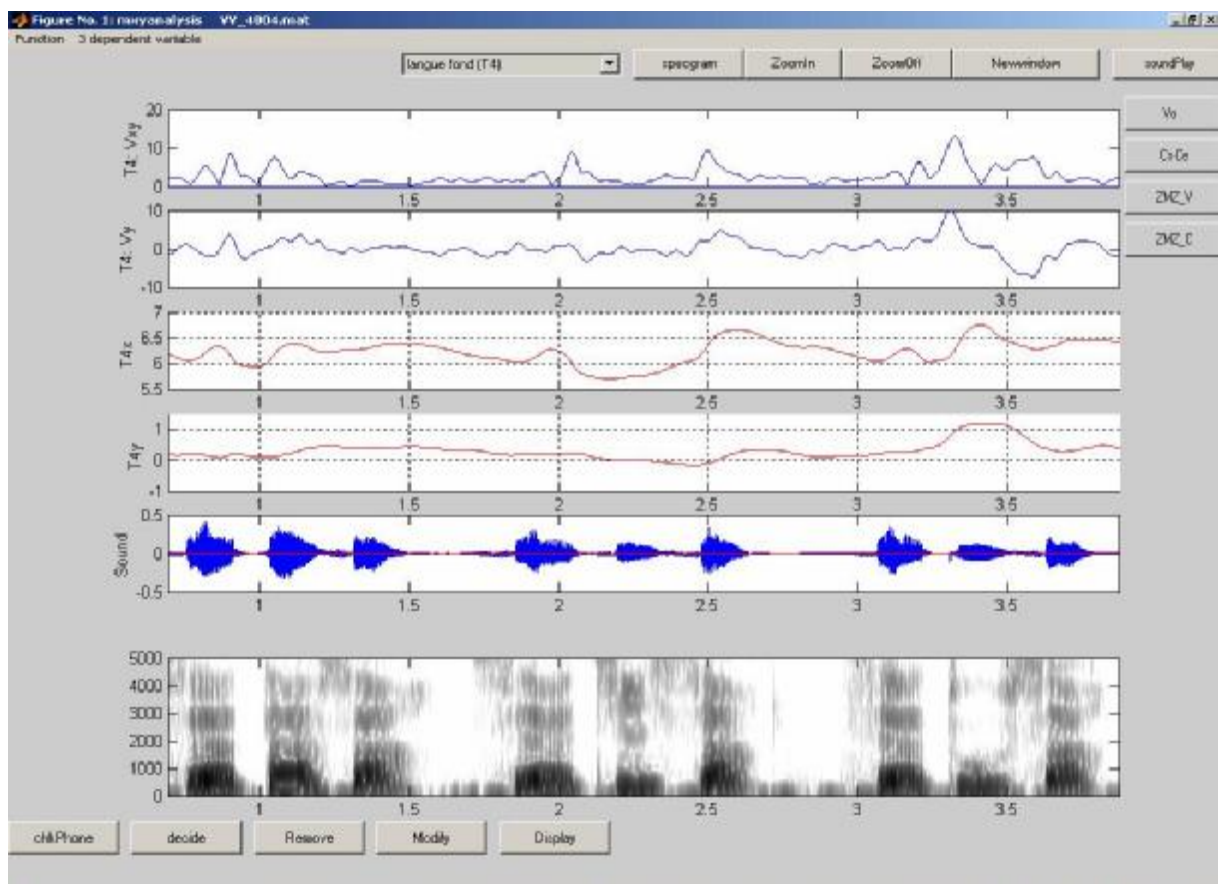


Figure 5.1 Interface d'analyse des données

En partant du haut : le premier tracé est la vitesse tangentielle du capteur T4, le deuxième tracé est la vitesse en Y, le troisième tracé est le déplacement en X, le quatrième tracé est le déplacement en Y, le cinquième tracé est le signal audio temporel, le sixième tracé est le spectrogramme du signal audio.

5.2.2 L'étiquetage acoustique

Le point caractéristique de l'articulation des phonèmes a d'abord été repéré manuellement sur les données acoustiques (Figure 5.2). La position de l'étiquette a été définie pour les voyelles /a, i, u/ en cherchant une partie stable sur le spectrogramme (stabilité maximale du patron des trois premiers formants). Pour les consonnes /t k/ on a choisi l'instant de début de l'explosion consonantique. Il s'agit d'un instant pertinent pour des consonnes plosives, car il

correspond selon Perrier *et al.* (2003) au point le plus proche de la cible intentionnelle de la consonne.

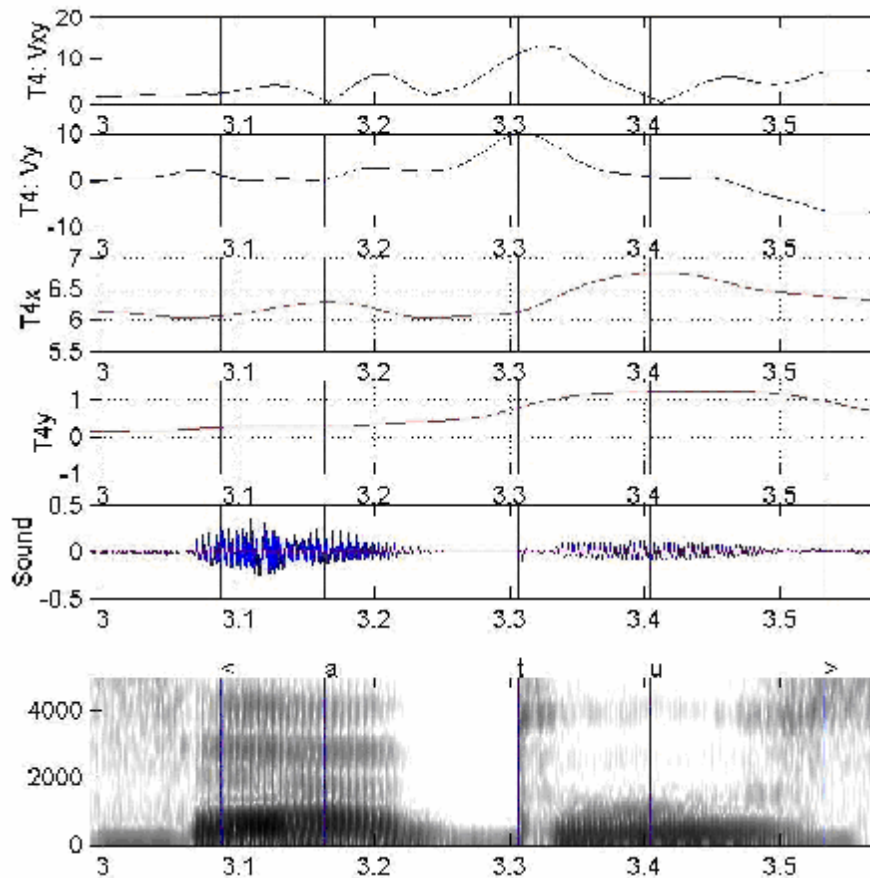


Figure 5.2 L'étiquetage acoustique

Même représentation que la figure 5.1. Les traits verticaux correspondent aux instants repérés pour les voyelles (/a/, /u/) et la consonne (/k/)

Pour l'étiquetage acoustique, nous nous concentrons sur le spectrogramme. Les symboles “<” et “>” sont utilisés afin de séparer la phrase porteuse de la partie que l'on étudie.

C'est <VCV> ça ? (En français)

这是 <VCV> 吗 ? (En chinois)

5.2.3 Modification automatique de l'étiquetage

L'étiquetage acoustique a été fait manuellement. Le critère était de choisir le point de stabilité maximale des trois premiers formants sur le spectrogramme. Ce point peut nous aider à bien trouver où se situe la voyelle et la consonne acoustiquement, mais il n'est pas toujours le point le plus représentatif du son dans l'espace articulatoire.

Afin de réaliser une détection plus précise de la configuration la plus canonique du conduit vocal, l'étiquette acoustique a été automatiquement modifiée vers la position la plus extrême du dos de la langue (T_4) autour de la position détectée sur le spectrogramme. La position la plus extrême du dos de la langue est définie pour chaque voyelle. Pour la voyelle /a/, nous avons choisi le point le plus bas du dos de la langue, pour la voyelle /u/, le point le plus en arrière du dos de la langue et pour /i/, le point le plus avant.

Pour permettre la recherche automatique de cette position canonique, nous avons observé la trajectoire de la langue pendant 100 ms ([-50ms, 50ms]) autour de l'étiquetage acoustique. T_4 a donc été pris en considération ici pour définir l'étiquetage articulaire, puisque nous avons considéré que c'était le meilleur index du positionnement global de la langue pour les voyelles.

La figure 5.3 illustre ce qui a été fait pour la voyelle /a/. Elle montre les trajectoires des 4 capteurs de la langue pendant 100ms autour de l'instant repéré sur le signal acoustique au centre de la zone de stabilité du patron formantique. Les 4 points ronds représentent la position des capteurs à cet instant. Les 4 points en étoile représentent la position de ces capteurs pour la configuration canonique. On voit que le capteur T_4 a atteint la position la plus basse de cette position de trajectoire. L'étiquetage articulaire ne concerne que les voyelles. Il ne nous est pas apparu nécessaire de modifier l'étiquetage des consonnes sur la base de données articulatoires.

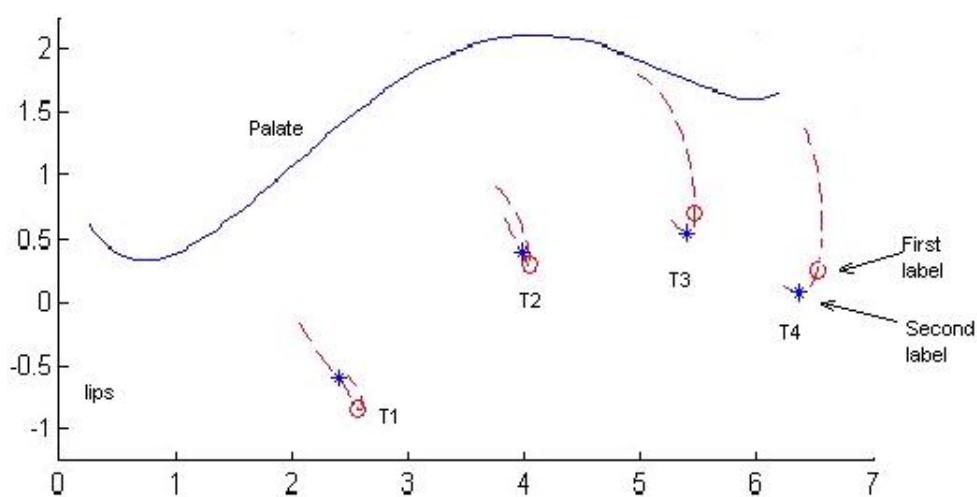


Figure 5.3 L'étiquetage articulaire

Le figure représente la trajectoire de la voyelle /a/ pour un intervalle de [-50ms 50ms] autour de l'étiquetage acoustique. Les points ronds donnent la position des capteurs à l'instant de l'étiquetage manuel fondé sur le signal acoustique. Les points en étoile représentent la position de ces capteurs retenue pour l'étiquetage articulaire. Le contour de palais est représenté en haut de la figure.

5.3 Enregistrement et modification du contour du palais

Pour bien analyser et présenter les données articulatoires, il est utile de considérer les positions des capteurs par rapport au palais. Pour cela, le contour du palais de chaque sujet a été enregistré. Un capteur a été placé sur le doigt, et le doigt a été déplacé contre le palais à l'intérieur de la bouche (de la glotte vers les dents). Le système EMMA a de cette façon enregistré le mouvement du doigt qui suit le contour du palais.

Même dans le cas où le déplacement du doigt est parfaitement resté dans le plan médian du palais, cette méthode génère un biais dans le positionnement du contour palatal.

Pour éliminer ce biais, nous avons modélisé le déplacement du capteur. Après avoir lissé le contour détecté par le passage du doigt le long du palais, nous avons déterminé les droites orthogonales à ce contour avec un pas d'un millimètre (cf, Figure 5.4). Puis en assimilant le capteur à un cercle de rayon 0.18 cm glissant sur le contour palatal, et en supposant que le point mesuré correspond au centre du capteur, nous avons décrit une meilleure description du palais en déplaçant le point mesuré de 0.18 cm vers l'extérieur du conduit vocal, le long de chacune de ces droites orthogonales (cf, Figure 5.4).

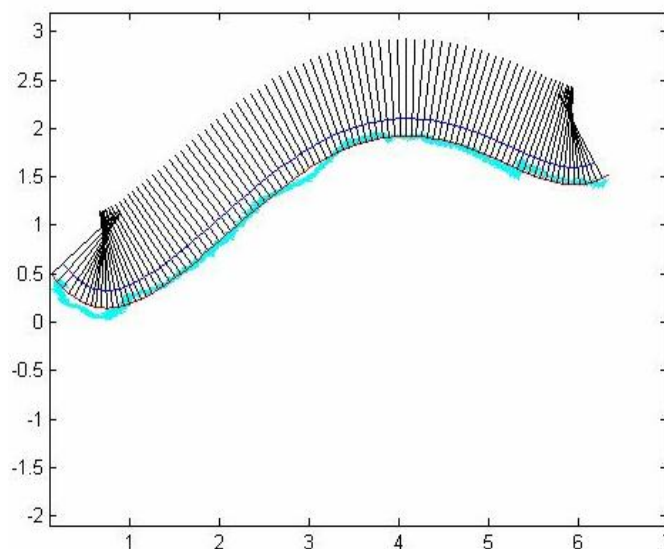


Figure 5.4 Contour du palais mesuré filtré et corrigé pour le sujet AV.

La courbe épaisse en bas est le contour mesuré. Le contour filtré est superposé en trait fin sur le contour mesuré. Le contour corrigé (en haut) a été obtenu en déplaçant le contour filtré de 0.18 cm vers l'extérieur le long de chacune des perpendiculaires. Lèvres à gauche –glotte à droite.

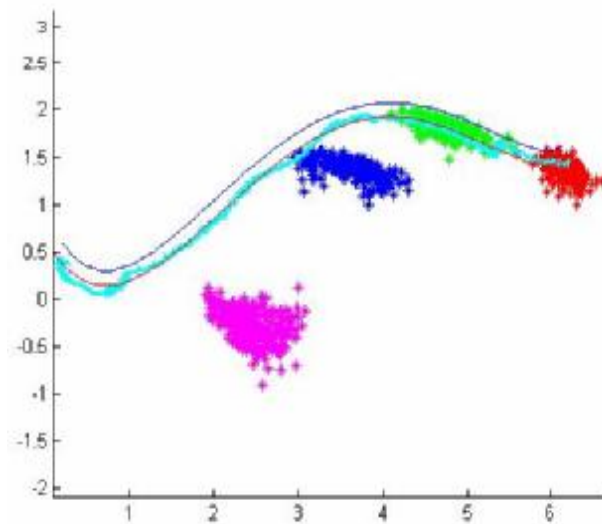


Figure 5.5 Contour du palais mesuré filtré, palais corrigé et données de la langue pour le sujet AV. Les 4 nuages de points représentent les 4 capteurs (T₁, T₂, T₃ et T₄) pour la consonne /k/ dans toutes les séquences V₁kV₂. Lèvres à gauche –glotte à droite.

La distance choisie pour déplacer le palais a été définie en vérifiant le contact avec le palais pour la consonne vélaire /k/. Nous avons analysé la consonne /k/ dans toutes les séquences VCV pour observer les contacts vélaire de la langue avec le palais. La figure 5.5 montre le palais mesuré filtré, le palais corrigé et les données de la langue de /k/ pour le sujet AV. Les 4 nuages de points représentent les 4 capteurs (T₁, T₂, T₃ et T₄, de gauche à droite). Nous avons observé que les capteurs T₃ et T₄ réalisent le contact dans la partie vélaire du conduit vocal. Mais nous constatons que le contact vélaire n'est pas cohérent avec le palais mesuré filtré. Nous avons déplacé ce palais mesuré filtré comme nous avons expliqué dans la figure 5.4, à la position de contact vélaire des données.

Nous avons fait le déplacement du palais pour tous les 6 sujets. La valeur moyenne de ce déplacement est 0.18cm. La figure 5.6 présente les palais des 6 sujets. Les palais mesurés sont à gauche, les palais corrigés sont à droite. Les nouveaux palais seront utilisés pour présenter nos résultats d'analyse des données articulatoires (chapitre 6).

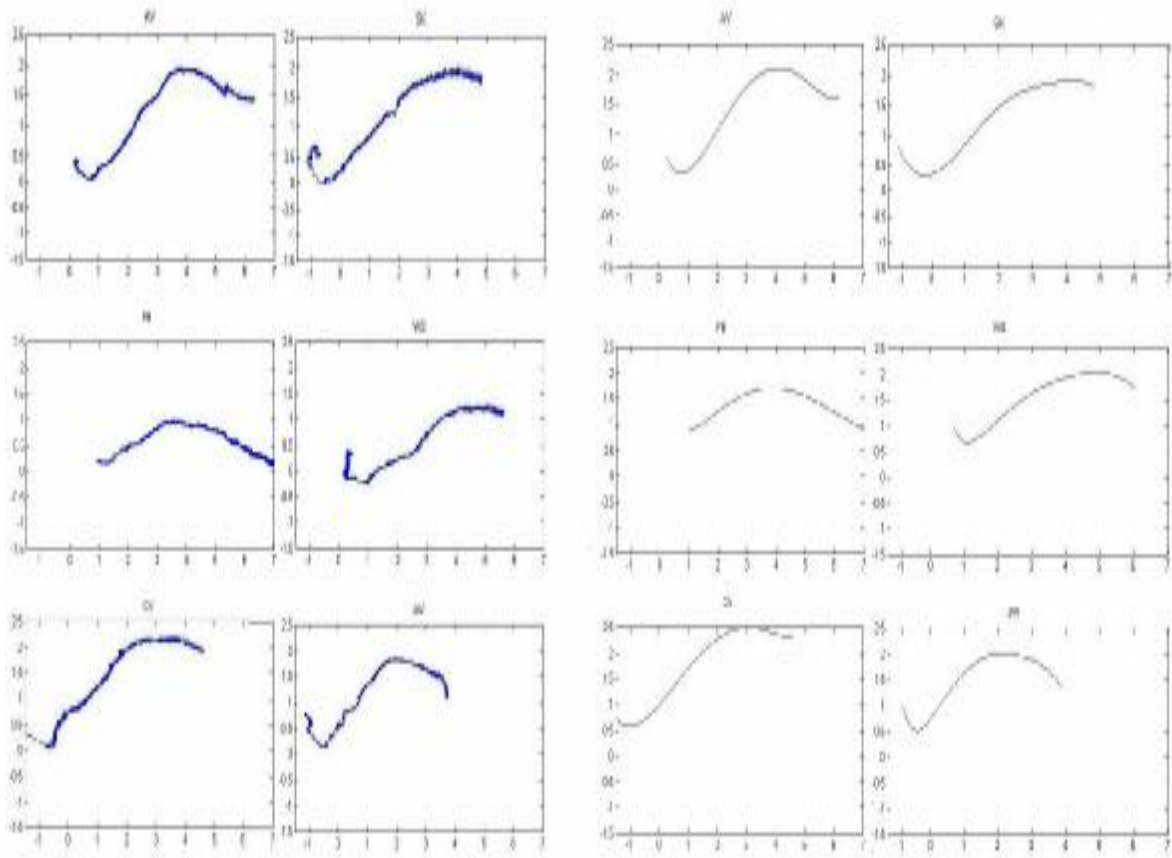


Figure 5.6 Les contours du palais des 6 sujets.

Les palais mesurés sont à gauche, les palais corrigés à droite. Pour chaque côté, la première colonne représente les trois sujets français (AV, PB et CV) et la deuxième colonne représente les trois sujets chinois (SK, WS et JW). Lèvres à gauche - glotte à droite.

Troisième partie : Résultats

Chapitre 6 : Analyse des données articulatoires

Introduction

Dans ce chapitre, nous présentons les analyses de la coarticulation anticipatoire pour la séquence V_1CV_2 en considérant l'influence de la voyelle V_2 sur la voyelle précédente V_1 (anticipation non adjacente) et sur la consonne C (anticipation non adjacente). Nous comparons les stratégies d'anticipation des différents sujets dans deux langues, le français et le chinois. Toutes les données sont analysées sur la position atteinte par langue pour V_1 et pour C .

Nous avons fait des analyses de variance de type à Mesures Répétées où les variables dépendantes étaient les positions horizontales et verticales des capteurs T_1 , T_2 , T_3 et T_4 à l'instant d'étiquetage, et la variable indépendante était la nature de la voyelle V_2 (/a/, /i/, /u/). Toutes les analyses ont été faites séparément pour les séquences $/V_1tV_2/$ et les séquences $/V_1kV_2/$, et pour chaque sujet. Nous avons comparé les comportements des sujets par groupe de langue. Les tests post-hoc (LSD) ont été effectués après les analyses de mesures répétées pour chaque sujet. Ils comparent toutes les moyennes de la position de la langue deux à deux par sujet. Nous présenterons pour les tests post-hoc seulement les résultats significatifs ($p < 0.05$). Le logiciel SPSS pour Windows a été utilisé pour cette analyse.

6.1 Anticipation non adjacente : Effet de la voyelle V_2 sur la voyelle V_1

La position atteinte par la langue pour les trois voyelles V_1 (/a/ /i/ /u/) a été donc analysée séparément en contexte consonantique /t/ et /k/. Dans chaque type de séquence (V_1 -alvéolaire- V_2 et V_1 -vélaire- V_2), lorsqu'on fixe la voyelle V_1 et que l'on fait varier la voyelle V_2 , nous observons des changements dans la forme de la langue pour V_1 .

La position de la langue est caractérisée par les trois capteurs T_2 , T_3 et T_4 lorsqu'ils sont disponibles. Les capteurs T_2 et T_3 situés sur la langue pour le sujet WS, le capteur T_3 du sujet AV, ainsi le capteur T_1 du sujet SK et du sujet CV n'ont pas fonctionné pendant la totalité de l'expérience (rupture du câble à la jonction avec le capteur). Nous ne les avons pas considérés dans notre analyse.

6.1.1 Position atteinte par la langue pour $V_1=/a/$ dans les séquences /ata/, /ati/ et /atu/

a. Sujets français

La figure 6.1 représente les positions de la langue pour $V_1 = /a/$ dans les séquences /ata/, /ati/ et /atu/ pour les trois sujets français PB, CV et AV. La figure montre, pour chaque sujet, l'ellipse de dispersion calculée à partir des mesures de toutes les répétitions. Le tableau 6.1 montre les résultats d'analyses ANOVA en fonction des contextes /ata/, /ati/ et /atu/. Le tableau 6.2 montre les résultats de tests post-hoc. Les différentes moyennes des positions de la langue (en cm) observée pour $V_1=/a/$ sont présentées seulement dans les cas statistiquement significatifs ($p<0.05$).

Observations :

- Le sujet PB montre des différences significatives pour la position verticale de T_2 et de T_3 . Les capteurs T_2 et T_3 sont plus bas pour $V_2=/a/$ que pour $V_2=/i/$ et pour $V_2=/u/$. La différence moyenne pour T_{2y} est (/atu/-/ata/) = 1.9mm et (/ati/-/ata/) = 1.6mm, et pour T_{3y} est (/atu/-/ata/) = 1.6mm et (/ati/-/ata/) = 1.8mm. Il montre aussi une différence sur la dimension avant/arrière entre $V_2=/i/$ et $V_2=/u/$. Les capteurs T_2 , T_3 et T_4 sont plus antérieurs pour $V_2=/i/$ que pour $V_2=/u/$.
- Le sujet CV montre des différences significatives pour les positions de T_2 , de T_3 , et de T_4 . Ces capteurs sont plus hauts pour $V_2=/i/$ que pour $V_2=/a/$ et pour $V_2=/u/$. La différence moyenne pour T_{2y} est (/ati/-/ata/) = 2.1mm et (/ati/-/atu/) = 1.7mm, pour T_{3y} est (/ati/-/ata/) = 2.6mm et (/ati/-/atu/) = 2.0mm, et pour T_{4y} est (/ati/-/ata/) = 1.9mm et (/ati/-/atu/) = 1.8mm;
- Le sujet AV montre une position de T_2 significativement plus haute pour $V_2=/i/$ que pour $V_2=/a/$ ou $V_2=/u/$. La différence moyenne en position verticale pour T_{2y} est (/ati/-/ata/) = 2.4mm et (/ati/-/atu/) = 1.6mm.
- Les trois sujets français montrent un effet d'anticipation significatif pour la dimension hauts/basse de la langue. La langue est plus basse quand la voyelle suivante (V_2) est une voyelle basse /a/ et est plus haute quand la voyelle suivante (V_2) est une voyelle haute /i/.

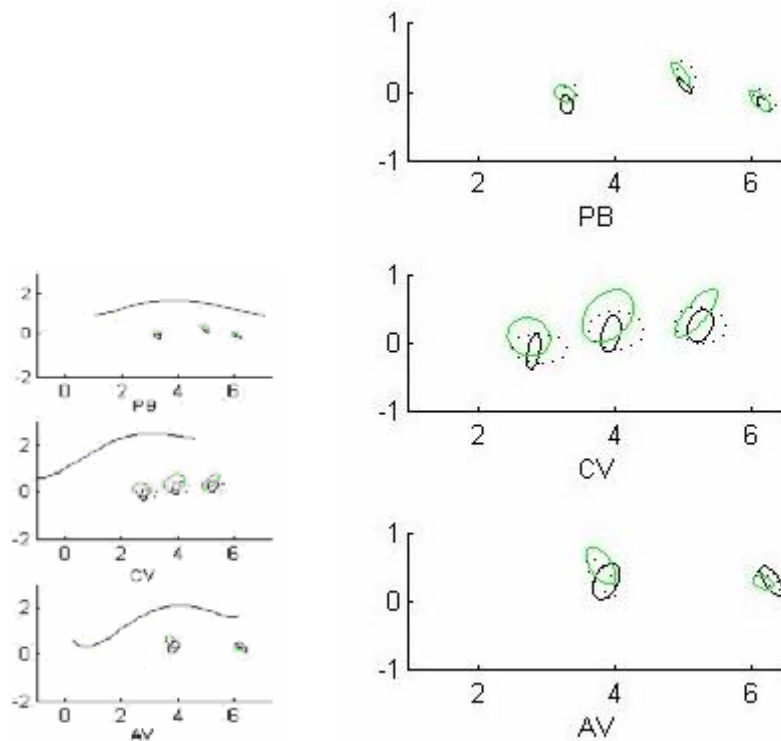


Figure 6.1 : Positions des capteurs (en cm) pour la voyelle $V_1 = /a/$ dans les séquences /ata/, /ati/ et /atu/ pour les trois sujets français (en partant du haut : PB CV AV). La partie gauche donne une vue générale de tous les capteurs avec la forme du palais comme référence géométrique ; la partie droite fournit un agrandissement sur les capteurs T_2 , T_3 et T_4 (quand ils sont disponibles). Les trois traits différents représentent V_2 (le trait noir pour /a/, le trait vert pour /i/ et le trait pointillé pour /u/). Lèvres à gauche – glotte à droite

/ a / in /ata/ /ati/ /atu/	PB	CV	AV
T2x	F=7.366 P=0.005	F=2.657 p=0.105	F=4.798 p=0.021
T2y	F=32.889 P=0	F=13.942 p=0	F=21.53 p=0
T3x	F=7.641 p=0.005	F=1.258 p=0.315	---
T3y	F=22.746 p=0	F=20.094 p=0	---
T4x	F=8.047 p=0.004	F=6.534 p=0.598	F=11.343 p=0.001
T4y	F=2.160 p=0.148	F=16.417 p=0	F=1.101 p=0.354

Tableau 6.1 Résultats d'analyses ANOVA pour les sujets français.
Le symbole " - - - " est présenté quand le capteur n'est pas disponible. Le gris représente les cas significatifs ($p < 0.05$).

	PB			CV			AV		
	/atu/ vs /ata/	/ati/ vs /ata/	/atu/ vs /ati/	/atu/ vs /ata/	/ati/ vs /ata/	/atu/ vs /ati/	/atu/ vs /ata/	/ati/ vs /ata/	/atu/ vs /ati/
T2x			0,08					- 0,08	
T2y	0,19	0,16			0,21	0,17		0,24	- 0,16
T3x			0,09				---	---	---
T3y	0,16	0,18			0,26	0,20	---	---	---
T4x		-0,06	0,08					- 0,13	0,07
T4y					0,19	0,18			

Tableau 6.2 Résultats de tests post-hoc pour les trois sujets français

Différence moyenne des positions de la langue (en cm) pour V_1 dans les séquences $/V_1tV_2/$. Seuls les cas statistiquement significatifs sont présentés ($p < 0.05$). La valeur est positive quand la position de capteur dans la deuxième séquence est plus postérieure ou plus haute que dans la première séquence, et négative quand elle est plus antérieure ou basse. Le symbole “- - -” est représenté quand le capteur n’est pas disponible.

b. Sujets chinois

La figure 6.2 représente les positions de la langue de $V_1=/a/$ dans les séquences $/ata/$, $/ati/$ et $/atu/$ pour les trois sujets chinois. Les tableaux 6.3 et 6.4 montrent les résultats d’analyses ANOVA et les résultats de tests post-hoc.

Observations :

- Le sujet SK ne montre pas de différence significative associée à V_2 qui peut être systématiquement observée à travers les capteurs.
- Le sujet JW montre des différences significatives sur les positions de T_2 , de T_3 et de T_4 . Ces capteurs sont plus postérieurs pour $V_2=/i/$ que pour $V_2=/a/$ et pour $V_2=/u/$. La différence moyenne pour T_{2x} est $(/ati/-/ata/)=3.2\text{mm}$ et $(/ati/-/atu/)=2.3\text{mm}$. Pour T_{3x} , la différence est $(/ati/-/ata/)=2.4\text{mm}$ et $(/ati/-/atu/)=2\text{mm}$, et pour T_{4x} , la différence est $(/ati/-/ata/)=1.7\text{mm}$ et $(/ati/-/atu/)=1.5\text{mm}$.
- Le sujet WS montre des différences significatives associées à V_2 sur la position horizontale de T_4 . Le capteur T_4 est plus antérieur pour $V_2=/i/$ que pour $V_2=/a/$ et pour $V_2=/u/$. La différence moyenne pour T_{4x} est $(/ata/-/ati/)=2.1\text{mm}$ et $(/atu/-/ati/)=1.2\text{mm}$.
- Nous n’avons pas observé d’effet d’anticipation systématique pour les trois sujets chinois. Le sujet JW montre un effet lié à V_2 sur la dimension avant/arrière, mais il est en contradiction avec un effet d’anticipation. La langue est en effet plus en arrière quand la voyelle suivante (V_2) est la voyelle antérieure $/i/$.

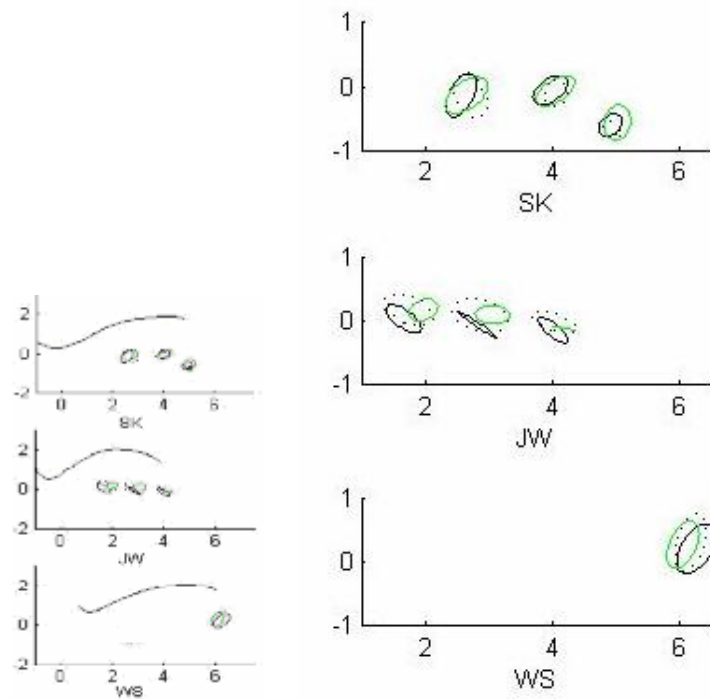


Figure 6.2 : Positions des capteurs (en cm) pour la voyelle $V_1 = /a/$ dans les séquences /ata/, /ati/ et /atu/ pour les trois sujets chinois (en partant du haut : SK, JW, WS). (Voir Figure 6.1 pour les détails)

/ a / in /ata/, /ati/, /atu/	SK	JW	WS
T2x	F=2.530 P=0.115	F=23.725 p=0.001	---
T2y	F=0.026 P=0.975	F=2.343 p=0.159	---
T3x	F=2.420 P=0.125	F=16.982 p=0.003	---
T3y	F=0.518 P=0.607	F=5.632 p=0.042	---
T4x	F=9.790 P=0.002	F=0.843 p=0.007	F=23.595 p=0
T4y	F=2.139 P=0.155	F=5.921 p=0.038	F=8.808 p=0.002

Tableau 6.3 Résultats d'analyses ANOVA pour les sujets chinois

	SK			JW			WS		
/a/	/atu/ vs /ata/	/ati/ vs /ata/	/atu/ vs /ati/	/atu/ vs /ata/	/ati/ vs /ata/	/atu/ vs /ati/	/atu/ vs /ata/	/ati/ vs /ata/	/atu/ vs /ati/
T2x					0,32	0,23	---	---	---
T2y							---	---	---
T3x					0,24	0,20	---	---	---
T3y				0,16	0,16		---	---	---
T4x		0,11			0,17	0,15		0,21	0,12
T4y				0,13			0,10		

Tableau 6.4 Résultats de tests post-hoc pour les trois sujets chinois

Remarque :

L'observation pour le sujet JW est contradictoire avec le fait que /i/ est articulée de façon plus en avant que les deux autres voyelles. Par conséquent, on ne peut pas considérer qu'il existe une anticipation directe de la voyelle suivante pour ce sujet. Nous supposons que c'est la conséquence de l'influence de la configuration articulatoire liée à la consonne / t /, quand elle est prononcée devant la voyelle /i/ par ce sujet. En effet, pour cette séquence, la langue du sujet JW pour la consonne /t/ (figure 6.3) est déjà localisée plus en arrière et le dos de la langue est plus haut et plus près du palais que devant une voyelle /a/ ou une voyelle /u/.

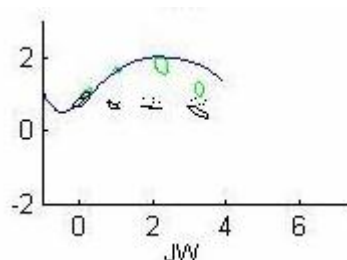


Figure 6.3 : Positions des capteurs (en cm) pour la consonne C = /t/ dans les séquences /ata/, /ati/ et /atu/ pour le sujet JW. La figure présente tous les capteurs avec la forme du palais comme référence géométrique. Les trois traits différents représentent V₂ (le trait noir pour /a/, le trait vert pour /i/ et le trait pointillé pour /u/). Lèvres à gauche –glotte à droite

Des explications possibles pour ce phénomène pourraient être trouvées à un niveau phonologique (/ti/ pourrait être plus affriquée que /tu/ ou /ta/) ou au niveau de la morphologie du palais (JW a notamment un palais plus arqué dans le plan sagittal que les deux autres sujets chinois SK et WS). Cependant, d'autres investigations sont nécessaires pour évaluer ces hypothèses. En effet, il semble raisonnable de considérer que la variabilité de V₁ associée aux changements de V₂ n'est pas le résultat direct d'une stratégie anticipée de la voyelle /i/.

6.1.2 Position atteinte par la langue pour $V_1=/i/$ dans les séquences /ita/, /iti/ et /itu/

a. Sujets français

La figure 6.4 représente les positions de la langue pour $V_1 = /i/$ dans les séquences /ita/, /iti/ et /itu/ pour les trois sujets français. Les tableaux 6.5 et 6.6 montrent les résultats d'analyses ANOVA et les résultats de tests post-hoc.

Observations :

- Le sujet PB montre une différence significative sur la position verticale de T_3 et de T_4 entre $V_2=/u/$ et les deux autres V_2 . T_3 est plus haut pour $V_2=/u/$ que pour $V_2=/a/$. T_4 est plus haut pour $V_2=/u/$ que pour $V_2=/i/$.
- Pour le sujet CV, une différence significative associée à V_2 est observée sur la position horizontale de T_2 et de T_3 . Ces deux capteurs sont plus en arrière pour $V_2=/u/$ que pour $V_2=/i/$ et/ou pour $V_2=/a/$. La différence moyenne pour T_{2x} est (/itu/-/iti/)=2.2mm ; pour T_{3x} est (/itu/-/ita/)=1.5mm et (/itu/-/iti/)=2.2mm.
- Le sujet AV montre une différence significative associée à V_2 sur la position horizontale de T_2 . La position de T_2 est plus en arrière pour $V_2=/u/$ que pour $V_2=/i/$ et pour $V_2=/a/$. Ce sujet montre aussi une différence sur T_4 , mais ce n'est pas un effet systématique pour les trois sujets.

b. Sujets chinois

La figure 6.5 représente les positions de la langue pour $V_1 = /i/$ dans les séquences /ita/, /iti/ et /itu/ pour les trois sujets chinois. Les tableaux 6.7 et 6.8 montrent les résultats d'analyses ANOVA et les résultats de tests post-hoc.

Observations :

- Le sujet SK montre seulement une différence significative sur la position horizontale de T_4 qui est plus antérieure pour $V_2=/i/$ que pour $V_2=/a/$ et pour $V_2=/u/$. La différence moyenne est (/itu/-/ita/)=0.9mm et (/iti/-/ita/)=1.4mm.
- Le sujet JW présente une différence significative associée à V_2 sur la position horizontale de T_4 . Le T_4 est plus antérieure pour $V_2=/a/$ que pour $V_2=/i/$ et pour $V_2=/u/$. La différence moyenne est (/ita/-/iti/)=1.4mm et (/itu/-/iti/)=1.3mm.
- Pour le sujet WS, une différence significative associée à V_2 est observée sur la position de T_4 . La position de T_4 est plus haute et plus en arrière pour $V_2=/u/$ que pour $V_2=/i/$.

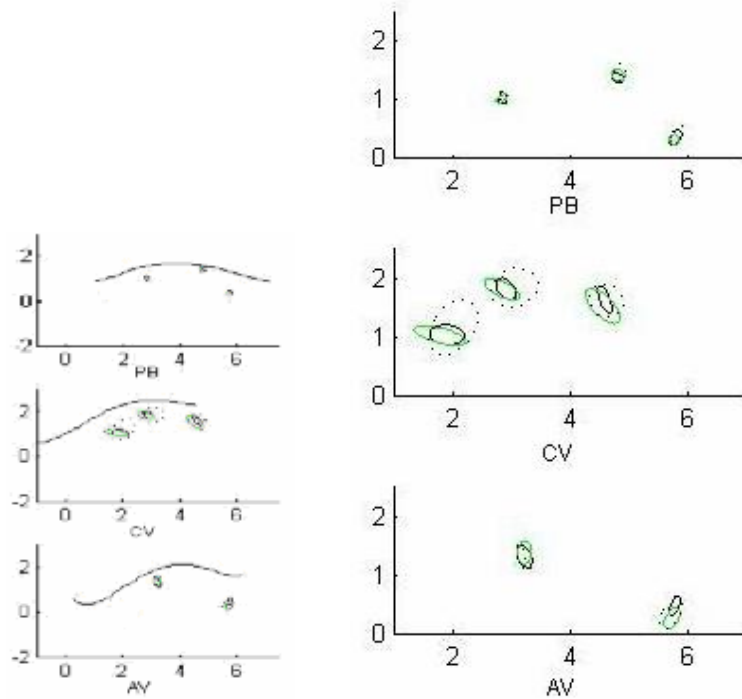


Figure 6.4 : Positions des capteurs (en cm) pour la voyelle $V_1 = /i/$ dans les séquences /ita/, /iti/ et /itu/ pour les sujets français.

$/i/$ in /ita/, /iti/, /itu/	PB	CV	AV
T2x	F=1.102 $p=0.354$	F=5.032 $p=0.020$	F=9.574 $p=0.001$
T2y	F=0.427 $p=0.659$	F=2.093 $p=0.156$	F=6.775 $p=0.006$
T3x	F=1.456 $p=0.259$	F=5.633 $p=0.014$	---
T3y	F=12.986 $p=0$	F=0.178 $p=0.839$	---
T4x	F=6.379 $p=0.008$	F=0.498 $p=0.617$	F=2.582 $p=0.103$
T4y	F=8.854 $p=0.002$	F=1.775 $p=0.201$	F=13.603 $p=0$

Tableau 6.5 Résultats d'analyses ANOVA pour les sujets français

	PB			CV			AV		
$/i/$	/itu/ vs /ita/	/iti/ vs /ita/	/itu/ vs /iti/	/itu/ vs /ita/	/iti/ vs /ita/	/itu/ vs /iti/	/itu/ vs /ita/	/iti/ vs /ita/	/itu/ vs /iti/
T2x						0,22	0,06		0,07
T2y								0,06	0,06
T3x				0,15		0,22	---	---	---
T3y	0,08						---	---	---
T4x			0,06						
T4y			0,10				0,14	0,20	

Tableau 6.6 Résultats de tests post-hoc pour les sujets français

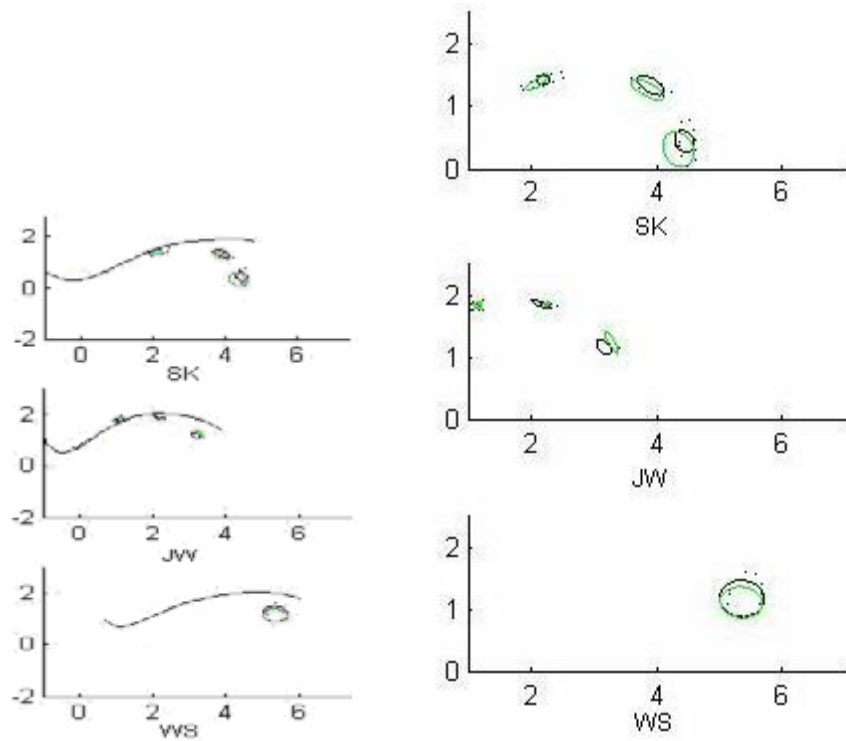


Figure 6.5 : Positions des capteurs (en cm) pour la voyelle $V_1 = /i/$ dans les séquences /ita/, /iti/ et /itu/ pour les sujets chinois.

/i/ in /ita/, /iti/, /itu/	SK	JW	WS
T2x	F=1.004 $p=0.388$	F=5.113 $p=0.079$	---
T2y	F=1.430 $p=0.268$	F=0.810 $p=0.507$	---
T3x	F=0.904 $p=0.425$	F=2.612 $p=0.188$	---
T3y	F=1.768 $p=0.203$	F=0.900 $p=0.476$	---
T4x	F=8.882 $p=0.003$	F=4.742 $p=0.038$	F=4.243 $p=0.031$
T4y	F=4.985 $p=0.170$	F=0.587 $p=0.598$	F=3.348 $P=0.048$

Tableau 6.7 Résultats d'analyses ANOVA pour les sujets chinois

	SK			JW			WS		
/i/	/itu/ vs /ita/	/iti/ vs /ita/	/itu/ vs /iti/	/itu/ vs /ita/	/iti/ vs /ita/	/itu/ vs /iti/	/itu/ vs /ita/	/iti/ vs /ita/	/itu/ vs /iti/
T2x							---	---	---
T2y							---	---	---
T3x							---	---	---
T3y							---	---	---
T4x		0,09	0,14	0,14	0,13				0,06
T4y									0,11

Tableau 6.8 Résultats de tests post-hoc pour les sujets chinois

6.1.3 Position atteinte par la langue pour $V_1=/u/$ dans les séquences /uta/, /uti/ et /utu/

a. Sujets français

La figure 6.6 représente les positions de la langue de $V_1 = /u/$ dans les séquences /uta/, /uti/ et /utu/ pour les trois sujets français. Les tableaux 6.9 et 6.10 montrent les résultats d'analyses ANOVA et les résultats de tests post-hoc.

Observations :

- Le sujet PB montre une différence significative sur la position verticale de T_2 et de T_3 . T_2 et T_3 sont plus hauts pour $V_2=/i/$ que pour $V_2=/a/$ ou pour $V_2=/u/$. La différence moyenne pour T_{2y} est (/uti/-/utu/) = 1.8mm et pour T_{3y} est (/uti/-/uta/) = 0.7mm.
- Pour le sujet CV, des différences significatives associées à V_2 existent en positions verticales de T_2 , T_3 . Ces capteurs sont plus hauts pour $V_2=/i/$ que pour $V_2=/a/$ et $V_2=/u/$. La différence moyenne pour T_{2y} est (/uti/-/uta/) = 3.3mm et (/uti/-/utu/) = 3.4mm. La différence moyenne pour T_{3y} est (/uti/-/uta/) = 2.6mm et (/uti/-/utu/) = 3.0mm.
- Le sujet AV montre une position de T_2 significativement plus haute pour $V_2=/i/$ que pour $V_2=/a/$ ou $V_2=/u/$. La différence moyenne pour T_{2y} est (/uti/-/uta/) = 1.3mm et (/uti/-/utu/) = 1.5mm.

b. Sujets chinois

La figure 6.7 représente les positions de la langue pour $V_1=/u/$ dans les séquences /uta/, /uti/ et /utu/ pour les trois sujets chinois. Les tableaux 6.11 et 6.12 montrent les résultats d'analyses ANOVA et les résultats de tests post-hoc.

Observations :

- Le sujet SK ne présente pas de différence significative en fonction de V_2 .
- Le sujet JW également, ne présente pas de différence significative en fonction de V_2 .
- Seule le sujet WS montre une différence sur T_{4x} . /utu/ est plus en arrière que /uti/. La différence moyenne est 1.5mm.

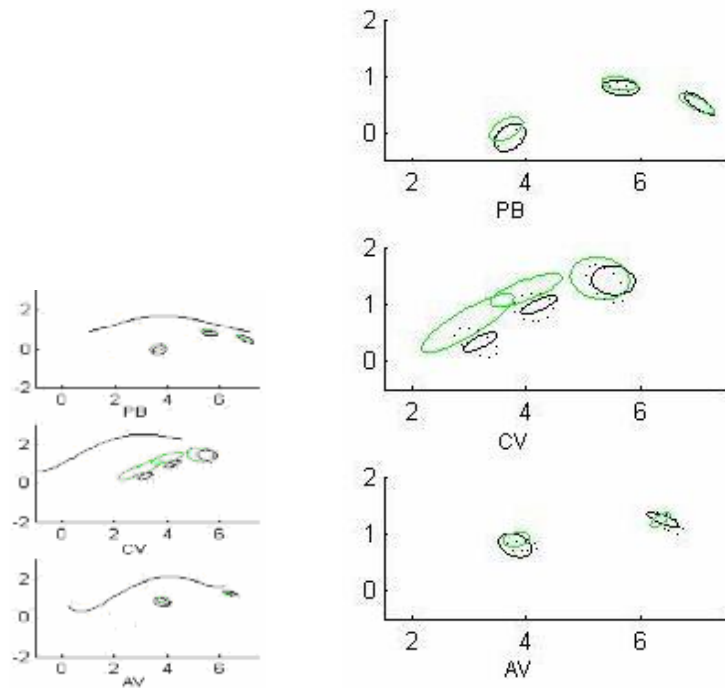


Figure 6.6: Positions des capteurs (en cm) pour la voyelle $V_1 = /u/$ dans les séquences /uta/, /uti/ et /utu/ pour les sujets français.

/ u / in /uta/,/uti/,/utu/	PB	CV	AV
T2x	F=0.719 $p=0.501$	F=1.803 $p=0.197$	F=4.842 $p=0.023$
T2y	F=8.123 $p=0.003$	F=15.555 $p=0$	F=11.678 $p=0.001$
T3x	F=0.101 $p=0.905$	F=2.335 $p=0.129$	---
T3y	F=4.959 $p=0.019$	F=22.910 $p=0$	---
T4x	F=0.657 $p=0.530$	F=4.861 $p=0.022$	F=6.357 $p=0.009$
T4y	F=0.144 $p=0.866$	F=2.933 $p=0.082$	F=1.427 $p=0.269$

Tableau 6.9 Résultats d'analyses ANOVA pour les sujets français

	PB			CV			AV		
/u/	/utu/ vs /uta/	/uti/ vs /uta/	/utu/ vs /uti/	/utu/ vs /uta/	/uti/ vs /uta/	/utu/ vs /uti/	/utu/ vs /uta/	/uti/ vs /uta/	/utu/ vs /uti/
T2x							0,14		0,11
T2y			- 0,18		0,33	- 0,34		0,13	-0,15
T3x							---	---	---
T3y		0,07			0,26	- 0,30	---	---	---
T4x					-0,24				0,15
T4y									

Tableau 6.10 Résultats de tests post-hoc pour les sujets français

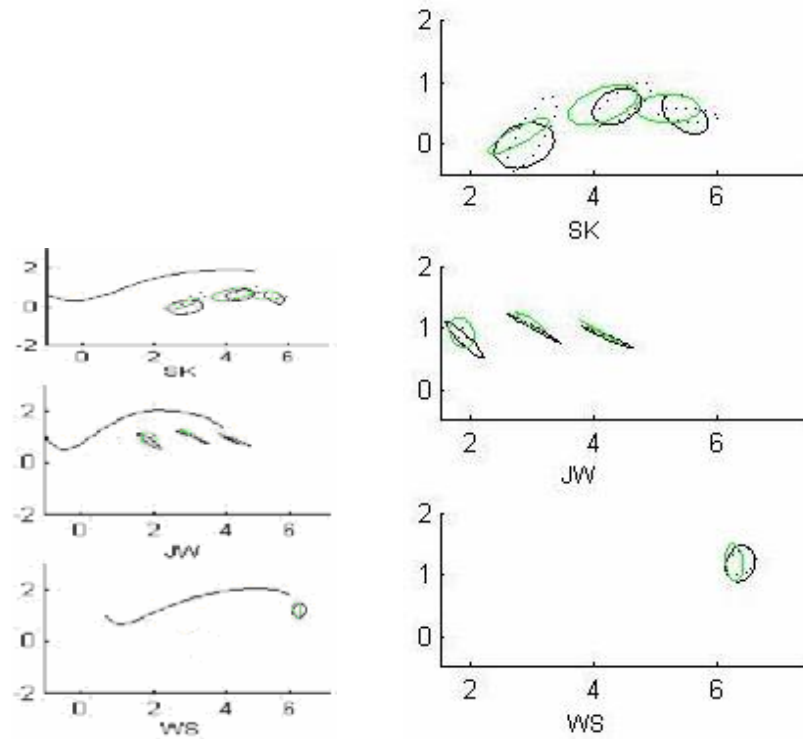


Figure 6.7 : Positions des capteurs (en cm) pour la voyelle $V_1 = /u/$ dans les séquences /uta/, /uti/ et /utu/ pour les sujets chinois.

/ u / in /uta/,/uti/,/utu/	SK	JW	WS
T2x	F=0.797 $p=0.483$	F=0.238 $p=0.799$	---
T2y	F=0.657 $p=0.544$	F=4.992 $p=0.082$	---
T3x	F=1.589 $p=0.262$	F=0.495 $p=0.643$	---
T3y	F=0.113 $p=0.894$	F=5.311 $p=0.075$	---
T4x	F=1.955 $p=0.204$	F=1.097 $p=0.417$	F=5.929 $p=0.012$
T4y	F=0.694 $p=0.527$	F=2.832 $p=0.171$	F=0.196 $p=0.824$

Tableau 6.11 Résultats d'analyses ANOVA pour les sujets chinois

	SK			JW			WS		
/u/	/utu/ vs /uta/	/uti/ vs /uta/	/utu/ vs /uti/	/utu/ vs /uta/	/uti/ vs /uta/	/utu/ vs /uti/	/utu/ vs /uta/	/uti/ vs /uta/	/utu/ vs /uti/
T2x									
T2y									
T3x									
T3y									
T4x									0,15
T4y									

Tableau 6.12 Résultats de tests post-hoc pour les sujets chinois

Résumé :

- Ü Première observation : La variabilité inter-répétition (représentée par la surface de l'ellipse) dépend fortement du sujet. Elle est très petite pour le sujet PB et le sujet JW, et très grande pour le sujet CV et le sujet SK, mais il y a une distribution similaire de cette variabilité pour les deux langues. Cette observation suggère que pour les voyelles /a/, /i/ et /u/ les contraintes motrices et les contraintes perceptives sur la variabilité inter-répétition sont similaires dans les deux langues.
- Ü Les sujets français montrent souvent une variabilité articulatoire significative pour V_1 quand V_2 change. La tendance principale de la variabilité mesurée soutient l'hypothèse selon laquelle en français l'articulation de V_1 dans les séquences V_1CV_2 anticipe l'articulation de V_2 . En effet, dans la majorité des cas, $V_2=/i/$ influence spécifiquement les positions T_2 et T_3 de V_1 , qui sont situées dans la région de constriction de /i/. Les positions plus hautes et/ou plus antérieures de T_2 et de T_3 de V_1 sont conformes à l'articulation suivante de /i/. Même dans de nombreux cas, la position T_2 et T_3 de V_1 montre des positions plus basses pour $V_2=/a/$ que pour $V_2=/i/$ et $V_2=/u/$. Ceci est conforme au fait que la langue, plus plate et plus basse, anticipe l'articulation de la prochaine voyelle, la voyelle basse /a/.
- Ü La variabilité de position de la langue associée au changement de V_2 est moins grande et moins systématique pour les sujets chinois que pour les sujets français. Pour une grande partie des cas, ils ne sont pas significatifs. Le sujet WS montre une différence sur la position horizontale de T_4 . Elle est plus antérieure pour $V_2=/i/$ que pour $V_2=/u/$ et/ou $V_2=/a/$. Une variation significative de V_1 pour JW dans les séquences /at V_2 / est observée, mais il ne semble pas que ce soit le résultat direct d'une stratégie anticipante de l'articulation suivante du /i/.
- Ü Pour les six sujets, la quantité de variabilité associée au changement de V_2 est fonction de la nature de la voyelle V_1 . La voyelle $V_1=/i/$ est en général moins sensible au changement de V_2 . Ceci est conforme au point de vue classique selon lequel /i/ est plus contrainte dans le domaine articulatoire que les autres voyelles (Recasens 1984, 1987).

6.1.4 Position atteinte par la langue pour $V_1=/a/$ dans les séquences /aka/ et /aku/

L'anticipation de V_2 sur la position atteinte par la langue pour V_1 dans les séquences V_1-k-V_2 est présentée dans cette section. Dans notre corpus, nous n'avons pas considéré la voyelle $V_2=/i/$ dans ce contexte pour respecter les règles phonologiques du chinois (/ki/ n'existe pas en chinois). Par conséquent, la différence de position de V_1 est associée seulement aux changements entre $V_2=/a/$ et $V_2=/u/$.

a. Sujets français

La figure 6.8 représente les positions de la langue de $V_1 = /a/$ dans les séquences /aka/ et /aku/ pour les trois sujets français. Le tableau 6.13 montre les résultats d'analyses ANOVA pour $V_1=/a/$ en fonction des contextes /aka/ et /aku/. Le tableau 6.14 montre les résultats de tests post-hoc. Les différences moyennes des positions de la langue (en cm) pour $V_1=/a/$ sont présentées seulement pour les cas statistiquement significatifs ($p<0.05$).

Observations :

- Le sujet PB montre une différence significative sur les positions verticales de T_2 et de T_4 . /aku/ est plus haute que /aka/ sur ces deux capteurs. La différence moyenne pour T_{2y} est (/aku/-/aka)=0.8mm et pour T_{4y} est (/aku/-/aka)=0.9mm.
- Pour le sujet CV, une différence significative associée à V_2 est observée sur la position verticale de T_2 . /aku/ est plus haute que /aka/ pour T_2 . La différence moyenne pour T_{2y} est (/aku/-/aka)=1.5mm.
- Le sujet AV ne présente pas de différence significative en fonction de V_2 .

b. Sujets chinois

La figure 6.9 représente les positions de la langue de $V_1 = /a/$ dans les séquences /aka/ et /aku/ pour les trois sujets chinois. Les tableaux 6.15 et 6.16 montrent les résultats d'analyses ANOVA et les résultats de tests post-hoc.

Observations :

- Le sujet SK ne présente pas d'effet significatif.
- Le sujet WS également, ne présente pas d'effet significatif.
- Le sujet JW montre une différence associée à V_2 sur les positions verticales de T_2 et de T_3 . /aku/ est plus haute que /aka/. La différence moyenne pour T_{2y} est (/aku/-/aka)=1.1mm et pour T_{3y} est (/aku/-/aka)=1.2mm.

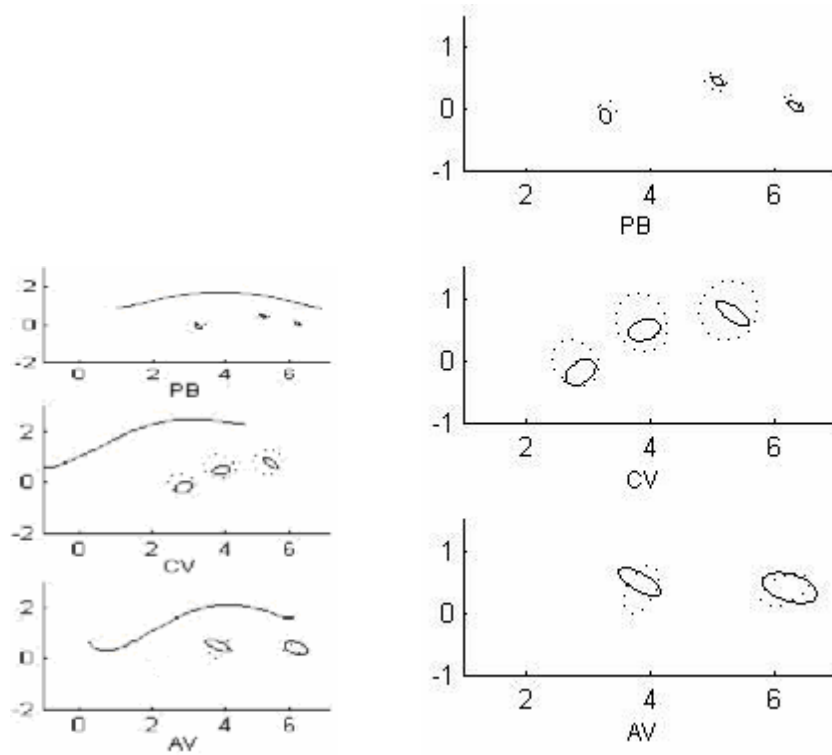


Figure 6.8 : Positions des capteurs (en cm) pour la voyelle $V_1 = /a/$ dans les séquences /aka/ et /aku/ pour les trois sujets français.

/ a / in /aka/,/aku/	PB	CV	AV
T2x	F=1.297 p=0.288	F=1.710 p=0.232	F=0.331 p=0.579
T2y	F=15.332 p=0.005	F=8.279 p=0.024	F=3.176 p=0.108
T3x	F=1.536 p=0.250	F=0.524 p=0.493	---
T3y	F=0.116 p=0.742	F=2.599 p=0.151	---
T4x	F=6.286 p=0.037	F=0.736 p=0.419	F=2.215 p=0.171
T4y	F=43.46 p=0	F=0.638 p=0.451	F=0.225 p=0.646

Tableau 6.13 Résultats d'analyses ANOVA pour les sujets français

	PB	CV	AV
/a/ vs /aku/ vs /aka/	/aku/ vs /aka/	/aku/ vs /aka/	/aku/ vs /aka/
T2y	0,08	0,15	
T3y			---
T4y	0,09		

Tableau 6.14 Résultats de tests post-hoc pour les trois sujets français

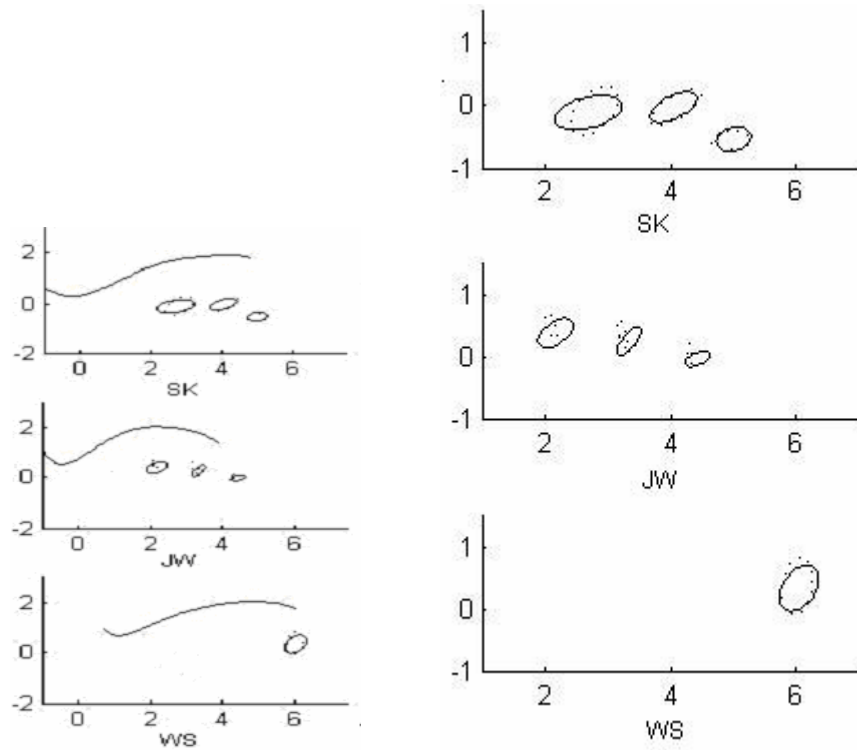


Figure 6.9 : Positions des capteurs (en cm) pour la voyelle $V_1 = /a/$ dans les séquences /aka/ et /aku/ pour les trois sujets chinois

/ a / in /aka/, /aku/	SK	JW	WS
T2x	F=1.467 p=0.265	F=0.925 p=0.438	---
T2y	F=0.733 P=0.420	F=289.0 p=0.003	---
T3x	F=0.338 p=0.579	F=2.385 p=0.263	---
T3y	F=0.036 p=0.854	F=108.0 p=0.009	---
T4x	F=0.908 p=0.372	F=3.990 p=0.184	F=0.230 P=0.644
T4y	F=0.427 p=0.535	F=7.839 p=0.107	F=8.000 p=0.570

Tableau 6.15 Résultats d'analyses ANOVA pour les sujets chinois

	SK	JW	WS
	/aku/ vs /aka/	/aku/ vs /aka/	/aku/ vs /aka/
T2y		0,11	---
T3y		0,12	---
T4y			

Tableau 6.16 Résultats de tests post-hoc pour les sujets chinois

6.1.5 Position atteinte par la langue pour $V_1=i/$ dans les séquences /ika/ et /iku/

a. Sujets français

La figure 6.10 représente les positions de la langue pour $V_1 = /i/$ dans les séquences /ika/ et /iku/ pour les trois sujets français. Les tableaux 6.17 et 6.18 montrent les résultats d'analyses ANOVA et les résultats de tests post-hoc.

Observations :

- Le sujet PB montre une différence en fonction de V_2 sur le capteur T_4 . /iku/ est plus en arrière et plus haute que /ika/. La différence moyenne pour T_{4x} est 1.1mm et pour T_{4y} est 0.8mm.
- Le sujet CV ne présente pas d'effet significatif.
- Le sujet AV ne présente pas non plus d'effet significatif.

b. Sujets chinois

La figure 6.11 représente les positions de la langue pour $V_1 = /i/$ dans les séquences /ika/ et /iku/ pour les trois sujets chinois. Les tableaux 6.19 et 6.20 montrent les résultats d'analyses ANOVA et les résultats de tests post-hoc.

Observations :

- Le sujet JW montre une différence associée à V_2 . /ika/ est plus en arrière que /iku/ sur T_2 . La différence est 0.8 mm. /ika/ est plus basse que /iku/ sur T_4 . La différence moyenne pour T_{4y} est 1.2mm.
- Le sujet SK ne présente pas d'effet significatif.
- Le sujet WS également, ne présente pas d'effet significatif.

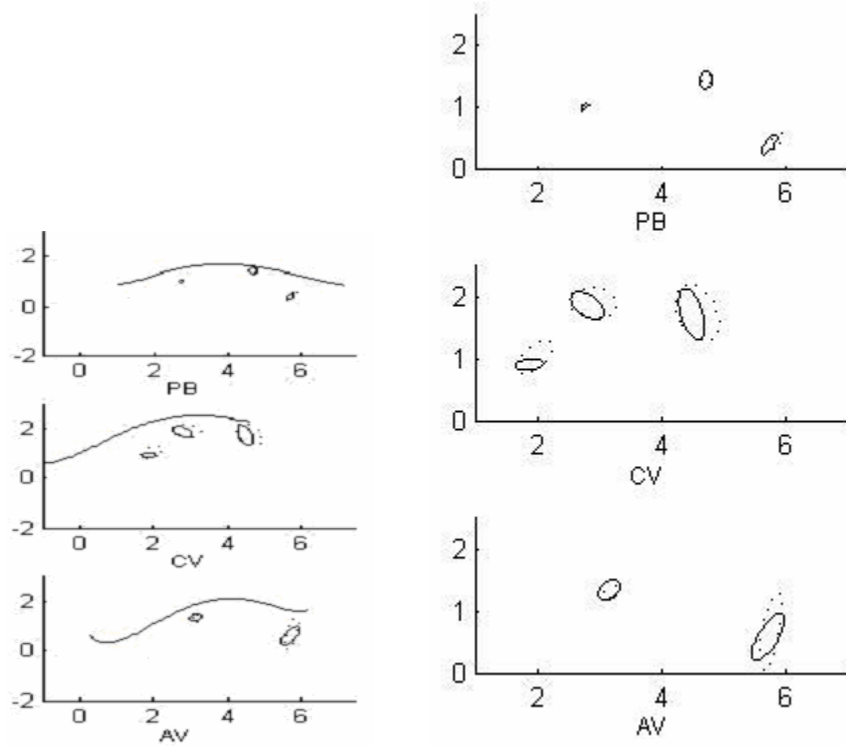


Figure 6.10 : Positions des capteurs (en cm) pour la voyelle $V_1 = /i/$ dans les séquences /ika/ et /iku/ pour les trois sujets français

/i/ in /ika/,/iku/	PB	CV	AV
T2x	F=1.332 $p=0.202$	F=3.072 $p=0.118$	F=0.067 $p=0.801$
T2y	F=1.606 $p=0.237$	F=4.407 $p=0.069$	F=0.199 $p=0.666$
T3x	F=1.187 $p=0.304$	F=2.644 $p=0.143$	---
T3y	F=0.024 $p=0.880$	F=0.466 $p=0.514$	---
T4x	F=41.029 $p=0$	F=1.956 $p=0.200$	F=0.543 $p=0.480$
T4y	F=4.932 $p=0.050$	F=0.083 $p=0.780$	F=0.533 $p=0.484$

Tableau 6.17 Résultats d'analyses ANOVA pour les sujets français

	PB	CV	AV
	/iku/ vs /ika/	/iku/ vs /ika/	/iku/ vs /ika/
T2x			
T2y			
T3x			---
T3y			---
T4x	0,11		
T4y	0,08		

Tableau 6.18 Résultats de tests post-hoc pour les sujets français

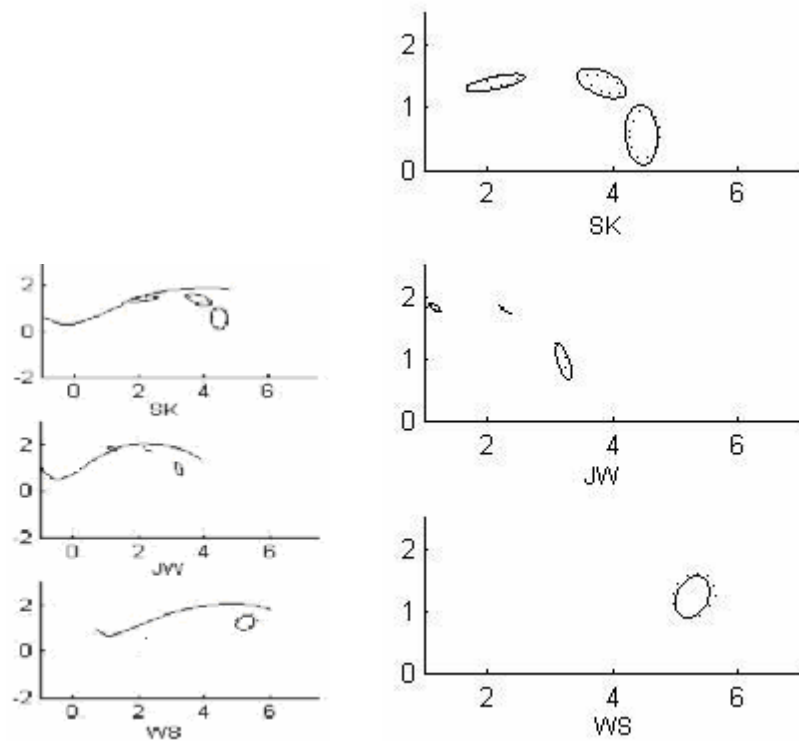


Figure 6.11 : Positions des capteurs (en cm) pour la voyelle $V_1 = /i/$ dans les séquences /ika/ et /iku/ pour les trois sujets chinois

/i/ in /ika/, /iku/	SK	JW	WS
T2x	F=0.836 $p=0.387$	F=156.25 $p=0.006$	---
T2y	F=0.911 $p=0.368$	F=0.429 $p=0.580$	---
T3x	F=0.135 $p=0.723$	F=1.286 $p=0.230$	---
T3y	F=0.013 $p=0.911$	F=6.75 $p=0.062$	---
T4x	F=0.696 $p=0.428$	F=12.00 $p=0.074$	F=0.428 $p=0.529$
T4y	F=0.018 $p=0.897$	F=48.00 $p=0.020$	F=0.118 $p=0.739$

Tableau 6.19 Résultats d'analyses ANOVA pour les sujets chinois

	SK	JW	WS
/i/ vs /ika/	/iku/ vs /ika/	/iku/ vs /ika/	/iku/ vs /ika/
T2x		0,08	---
T2y			---
T3x			---
T3y			---
T4x			
T4y		0,12	

Tableau 6.20 Résultats de tests post-hoc pour les sujets chinois

6.1.6 Position atteinte par la langue pour $V_1=/u/$ dans les séquences /uka/ et /uku/

a. Sujets français

La figure 6.12 représente les positions de la langue pour $V_1=/u/$ dans les séquences /uka/ et /uku/ pour les trois sujets français. Les tableaux 6.21 et 6.22 montrent les résultats d'analyses ANOVA et les résultats de tests post-hoc.

Observations :

- Pour le sujet PB, une différence significative en fonction de V_2 est observée sur les positions horizontales de T_3 et de T_4 . /uka/ est plus en arrière que /uku/ sur ces deux capteurs. La différence moyenne est 1.6mm.
- Le sujet CV montre une différence associée à V_2 sur T_2 . /uka/ est plus en arrière et plus bas que /uku/. La différence moyenne pour T_{2x} est 1.7mm et pour T_{2y} est 1.3mm.
- Pour le sujet AV, une différence sur les positions horizontales de T_2 et de T_4 est observée. /uka/ est plus en arrière que /uku/ sur ces deux capteurs. La différence moyenne pour T_{2x} est 2.4mm et pour T_{4x} est 2.7mm.

b. Sujets chinois

La figure 6.13 représente les positions de la langue pour $V_1 = /u/$ dans les séquences /uka/ et /uku/ pour les trois sujets chinois. Le tableau 6.23 montre les résultats d'analyses ANOVA.

Observations :

- Les trois sujets chinois ne montrent pas d'effet significatif

Résumé :

- Ü Les sujets français présentent à nouveau un effet d'anticipation de la voyelle V_2 dans la voyelle V_1 , mais il est plus petit que celui des séquences en contexte /t/. Les sujets chinois ne présentent aucun effet significatif.
- Ü Les résultats montrent que la quantité de variabilité de V_1 associée au changement de V_2 est non seulement fonction de la voyelle V_1 , mais également de la consonne C. La consonne /k/ impose plus de contraintes pour le mouvement du corps de langue, ce qui influence la coarticulation de voyelle à voyelle.

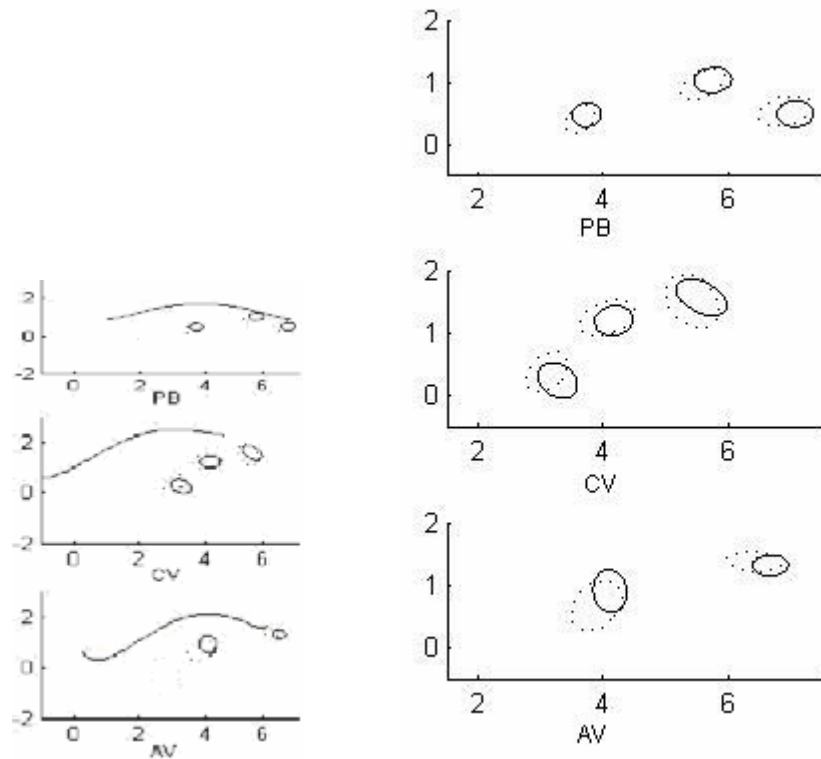


Figure 6.12 : Positions des capteurs (en cm) pour la voyelle $V_1 = /u/$ dans les séquences /uka/ et /uku/ pour les trois sujets français

/u/ in /uka/, /uku/	PB	CV	AV
T2x	F=4.139 $p=0.072$	F=8.038 $p=0.025$	F=12.778 $p=0.007$
T2y	F=4.079 $p=0.074$	F=9.135 $p=0.020$	F=25.051 $p=0.001$
T3x	F=5.920 $p=0.038$	F=2.419 $p=0.164$	---
T3y	F=4.047 $p=0.075$	F=0.251 $p=0.632$	---
T4x	F=6.446 $p=0.032$	F=4.954 $p=0.061$	F=20.505 $p=0.002$
T4y	F=1.073 $p=0.327$	F=1.199 $p=0.310$	F=2.980 $p=0.123$

Tableau 6.21 Résultats d'analyses ANOVA pour les sujets français

	PB	CV	AV
/u/	/uku/ vs /uka/	/uku/ vs /uka/	/uku/ vs /uka/
T2x		0,17	0,22
T2y		0,13	0,24
T3x	0,16		---
T3y			---
T4x	0,16		0,27
T4y			

Tableau 6.22 Résultats de tests post-hoc pour les sujets français

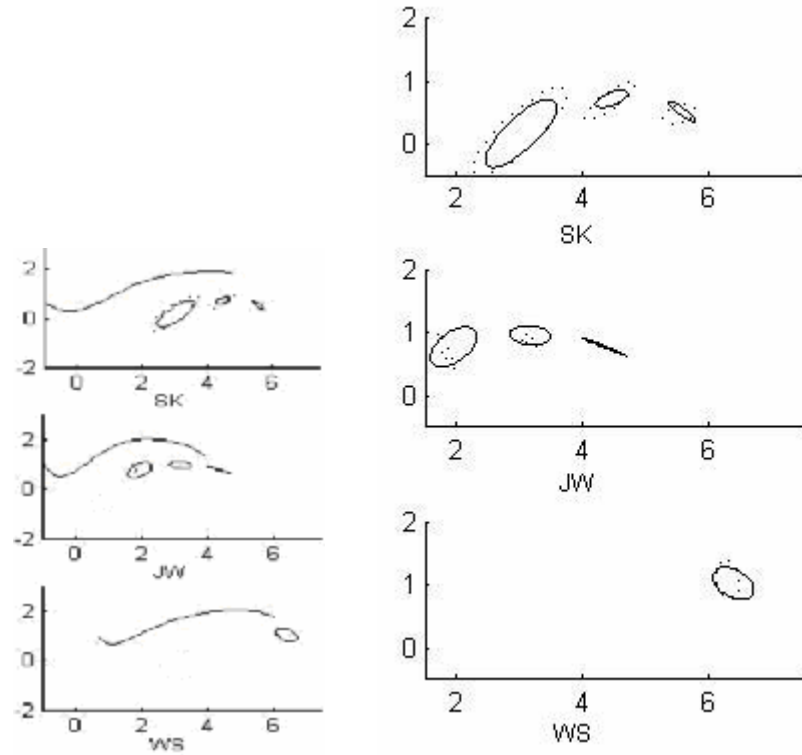


Figure 6.13 : Positions des capteurs (en cm) pour la voyelle $V_1 = /u/$ dans les séquences /uka/ et /uku/ pour les trois sujets chinois

/ u / in /uka/,/uku/	SK	JW	WS
T2x	F=0.090 p=0.776	F=1.551 p=0.339	---
T2y	F=0.060 p=0.816	F=1.497 p=0.346	---
T3x	F=0.128 p=0.735	F=2.326 p=0.267	---
T3y	F=0.277 p=0.621	F=1.750 p=0.317	---
T4x	F=0.611 p=0.470	F=8.308 p=0.102	F=2.237 P=0.169
T4y	F=0.649 p=0.457	F=5.297 p=0.148	F=3.676 P=0.087

Tableau 6.23 Résultats d'analyses ANOVA pour les sujets chinois

6.2 Anticipation adjacente : Effet de la voyelle V₂ sur C

Dans cette section, nous allons analyser la position atteinte par la langue pour la consonne C en contexte V₂. Lorsque nous fixons V₁C et que nous faisons varier V₂ parmi les voyelles étudiées (ex : /ata/ /atu/ /ati/), nous observons une modification des variations de la forme de la langue pour la consonne C. La position de la langue est caractérisée par les quatre capteurs T₁, T₂, T₃ et T₄ lorsqu'ils sont disponibles.

6.2.1 Position atteinte par la langue pour C=/t/ dans les séquences /ata/, /ati/ et /atu/

a. Sujets français

La figure 6.14 représente les positions de la langue pour C = /t/ dans les séquences /ata/, /ati/ et /atu/ pour les trois sujets français. Le tableau 6.25 montre les résultats d'analyses ANOVA pour C =/t/ en contextes /ata/, /ati/ et /atu/. Le tableau 6.26 montre les résultats de tests post-hoc. Les différences moyennes des positions de la langue (en cm) pour C=/t/ sont présentées seulement dans les cas statistiquement significatifs (p<0.05).

Observations :

- La différence principale et la plus remarquable associée à V₂ concernant la position verticale de T₂ et/ou de T₃ chez les trois sujets. Cette position est plus haute pour V₂=/i /, que pour V₂ = /u /, alors que la position est la plus basse pour V₂=/a/. La différence moyenne pour T_{2y} : (/ati/-/ata/) est 4.5mm pour PB, 3.7mm pour CV et 9.6mm pour AV ; (/ati/-/atu/) est 4.8mm pour PB, 3.4mm pour CV et 7.9mm pour AV ; pour T_{3y} : (/ati/-/ata/) est 5.7mm pour PB et 7.3mm pour CV ; (/ati/-/atu/) est 4.5mm pour PB et 5.0mm pour CV.
- Les trois sujets montrent une position verticale de T₄ plus haute pour V₂=/u/ que pour V₂=/a/ et pour V₂=/i/.
- Une différence associée à V₂ sur la direction horizontale est également observée, mais elle n'est pas systématique pour tous les capteurs et tous les sujets.

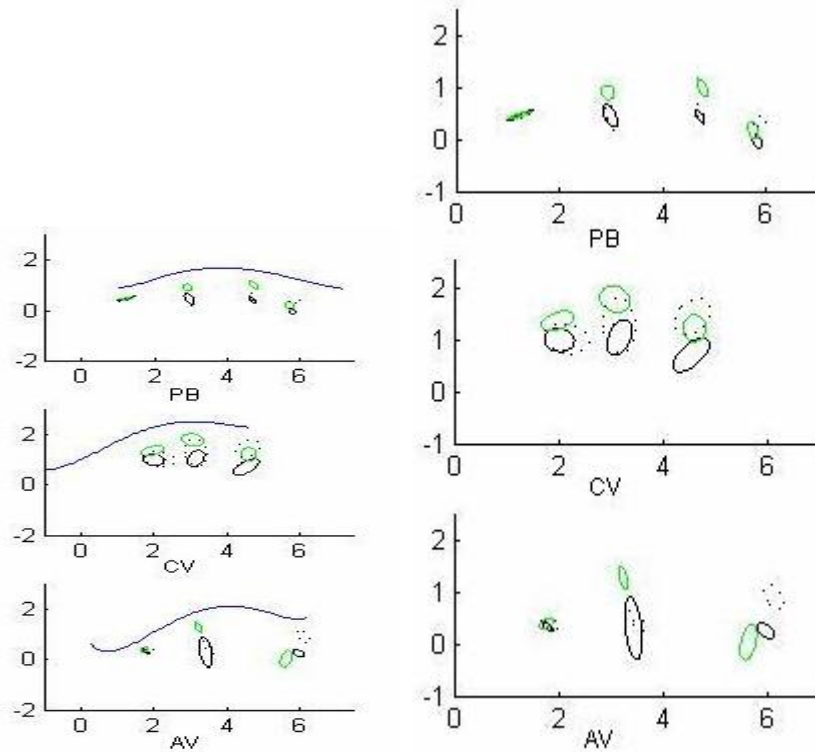


Figure 6.14: Positions des capteurs (en cm) pour la consonne C = /t/ dans les séquences /ata/, /ati/ et /atu/ pour les trois sujets français (en partant du haut : PB, CV et AV). La partie gauche donne une vue générale de tous les capteurs avec la forme de palais comme référence géométrique ; la partie droite fournit un agrandissement sur les capteurs T1, T2, T3 et T4 (quand ils sont disponibles). Les trois traits différents représentent V₂ (le trait noir pour /a/, le trait vert pour /i/ et le trait pointillé pour /u/). Lèvres à gauche –glotte à droite

/ t / in /ata/ /ati/ /atu/	PB	CV	AV
T1x	F=0.040 p=0.961	---	F=14.976 p=0
T1y	F=0.889 p=0.430	---	F=2.772 p=0.089
T2x	F=3.610 p=0.050	F=5.287 P=0.048	F=59.514 p=0
T2y	F=80.974 p=0	F=23.893 P=0	F=131.71 p=0
T3x	F=6.629 p=0.008	F=1.084 P=0.365	---
T3y	F=184.80 p=0	F=29.904 P=0	---
T4x	F=17.417 p=0	F=0.354 P=0.708	F=99.156 p=0
T4y	F=103.092 p=0	F=63.509 P=0	F=167.77 p=0

Tableau 6.25 Résultats d'analyses ANOVA pour les sujets français (PB, CV et AV) (C=/t/ dans les séquences /ata/, /ati/ et /atu/). Le gris représente les cas significatifs.

	PB			CV			AV		
/t/	/atu/vs /ata/	/ati/vs /ata/	/atu/vs /ati/	/atu/vs /ata/	/ati/vs /ata/	/atu/vs /ati/	/atu/vs /ata/	/ati/vs /ata/	/atu/vs /ati/
T1x				---	---	---	0,09		0,12
T1y				---	---	---			
T2x			0,06	0,14		0,17	0,06	-0,20	0,26
T2y		0,45	-0,48		0,37	-0,34	0,17	0,96	-0,79
T3x		0,05	-0,07				---	---	---
T3y	0,12	0,57	-0,45	0,23	0,73	-0,50	---	---	---
T4x		-0,07	0,11				0,15	-0,33	0,48
T4y	0,38	0,22	0,16	0,73	0,51	0,22	0,65	-0,23	0,87

Tableau 6.26 Résultats de tests post-hoc pour les sujets français

b. Sujets chinois

La figure 6.15 représente les positions de la langue pour C = /t/ dans les séquences /ata/, /ati/ et /atu/ pour les trois sujets chinois. Les tableaux 6.27 et 6.28 montrent les résultats d'analyses ANOVA et les résultats de tests post-hoc.

Observations :

- La variabilité de C associée à V₂ est très remarquable sur la dimension verticale pour T₂ et pour T₃ chez les sujets SK et JW. Cette position est plus haute pour V₂=i /, que pour V₂ = /u /, alors que la position la plus basse est pour V₂=a/. La différence moyenne pour T_{2y} : (/ati/-/ata/) est 7.7mm pour SK et 9.7mm pour JW; (/ati/-/atu/) est 6.8mm pour SK et 8.8mm pour JW ; pour T_{3y} : (/ati/-/ata/) est 10.4mm pour SK et 11.6mm pour JW; (/ati/-/atu/) est 5.3mm pour SK et 9.6mm pour JW.
- Les trois sujets montrent que la position verticale de T₄ est systématiquement plus haute pour V₂=i/ que pour V₂=u/, alors que la position est la plus basse pour V₂=a/. La différence moyenne pour T_{4y} : (/ati/-/ata/) est 10.9mm pour SK, 6.2mm pour JW et 7.4mm pour WS; (/ati/-/atu/) est 3.1mm pour SK et 3.2mm pour JW ; (/atu/-/ata/) est 7.8mm pour SK, 3.0mm pour JW et 6.6mm pour WS.
- Une différence dans la direction horizontale est également observée, mais elle n'est pas systématique pour tous les capteurs T₁, T₂, T₃, T₄ et pour les trois sujets.

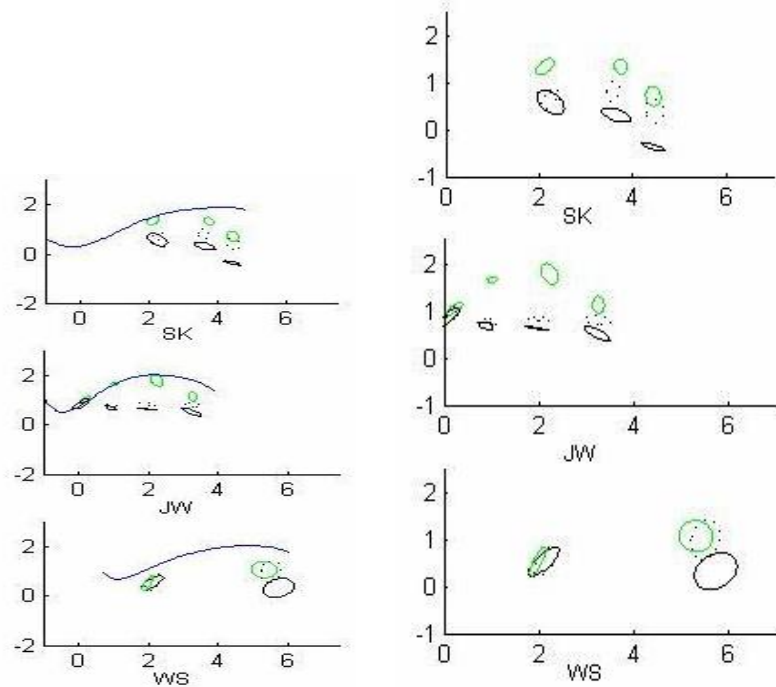


Figure 6.15 : Positions des capteurs (en cm) pour la consonne C = /t/ dans les séquences /ata/, /ati/ et /atu/ pour les trois sujets chinois

/ t / in /ata/, /ati/, /atu/	SK	JW	WS
T1x	---	F=7.946 p=0.021	F=14.39 p=0
T1y	---	F=18.592 p=0.003	F=0.767 p=0.479
T2x	F=4.841 p=0.025	F=3.626 p=0.093	---
T2y	F=104.91 p=0	F=1631.2 p=0	---
T3x	F=9.456 p=0.003	F=8.843 p=0.016	---
T3y	F=380.58 p=0	F=342.85 p=0	---
T4x	F=0.430 p=0.659	F=0.076 p=0.928	F=51.64 p=0
T4y	F=315.43 p=0	F=93.359 p=0	F=380.07 p=0

Tableau 6.27 Résultats d'analyses ANOVA pour les sujets français

	SK			JW			WS		
/t/	/atu/ vs /ata/	/ati/ vs /ata/	/atu/ vs /ati/	/atu/ vs /ata/	/ati/ vs /ata/	/atu/ vs /ati/	/atu/ vs /ata/	/ati/ vs /ata/	/atu/ vs /ati/
T1x	---	---	---		0,13	-0,10	0,06	-0,12	0,18
T1y	---	---	---		0,20	-0,17			
T2x		-0,12	0,10				---	---	---
T2y	0,09	0,77	-0,68	0,09	0,97	-0,88	---	---	---
T3x	-0,06	0,11	-0,17		0,28	-0,24	---	---	---
T3y	0,51	1,04	-0,53	0,20	1,16	-0,96	---	---	---
T4x							-0,22	-0,41	0,20
T4y	0,78	1,09	-0,31	0,30	0,62	-0,32	0,66	0,74	-0,08

Tableau 6.28 Résultats de tests post-hoc pour les sujets chinois

6.2.2 Position atteinte par la langue pour C=/t/ dans les séquences /ita/, /iti/ et /itu/

a. Sujets français

La figure 6.16 représente les positions de la langue pour C = /t/ dans les séquences /ita/, /iti/ et /itu/ pour les trois sujets français. Les tableaux 6.29 et 6.30 montrent les résultats d'analyses ANOVA et les résultats de tests post-hoc.

Observations :

- Nous avons observé le même phénomène pour la consonne /t/ en contexte $V_1=/i/$ qu'en contexte $V_1=/a/$ (figure 6.14). Les différences principales concernent la position verticale de T_2 et/ou de T_3 . Cette position est plus haute pour $V_2=/i/$, que pour $V_2=/u/$ et pour $V_2=/a/$. La différence moyenne pour T_{2y} : (/iti/-/ita/) est 3.4mm pour le sujet PB, 2.1mm pour le sujet CV et 4.9mm pour le sujet AV; (/iti/-/itu/) est 3.2mm pour le sujet PB, 1.1mm pour le sujet CV et 4.9mm pour le sujet AV. pour T_{3y} : (/iti/-/ita/) est 6.8mm pour le sujet PB et 5.6mm pour le sujet CV; (/iti/-/itu/) est 5.5mm et 3.8mm pour le sujet PB et le sujet CV.

b. Sujets chinois

La figure 6.17 représente les positions de la langue de C = /t/ dans les séquences /ita/, /iti/ et /itu/ pour les trois sujets chinois. Les tableaux 6.31 et 6.32 montrent les résultats d'analyses ANOVA et les résultats de tests post-hoc.

Observations :

- Les différences associées à V_2 sont très remarquables sur la position verticale pour T_2 , T_3 et T_4 . Cette position est plus haute pour $V_2=/i/$, que pour $V_2 = /u/$, alors que la position est plus basse pour $V_2=/a/$. La différence moyenne pour T_{2y} : (/iti/-/ita/) est 4.5mm pour SK et 9.5mm pour JW; (/iti/-/itu/) est 5.1mm pour SK et 7.9mm pour JW. La différence moyenne pour T_{3y} : (/iti/-/ita/) est 5.8mm pour SK et 10.9mm pour JW; (/iti/-/itu/) est 3.6mm pour SK et 8.6mm pour JW. pour T_{4y} : (/iti/-/ita/) est 5.2mm pour SK, 6.1mm pour JW et 7.2mm pour WS; (/itu/-/ita/) est 5.4mm pour SK, 2.3mm pour JW et 7.8mm pour WS.

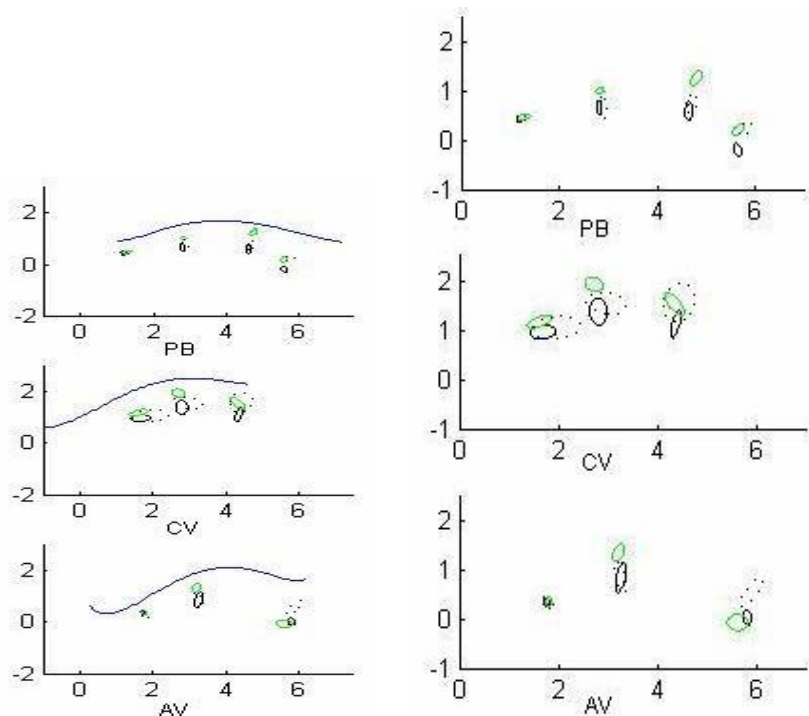


Figure 6.16 : Positions des capteurs (en cm) pour la consonne C = /t/ dans les séquences /ita/, /iti/ et /itu/ pour les trois sujets français

/t/ in /ita/, /iti/, /itu/	PB	CV	AV
T1x	F=9.931 p=0.001	---	F=7.220 p=0.005
T1y	F=6.515 p=0.007	---	F=3.116 p=0.069
T2x	F=22.571 p=0	F=20.556 p=0	F=2.894 p=0.081
T2y	F=60.519 p=0	F=12.656 p=0.001	F=86.435 p=0
T3x	F=50.969 p=0	F=11.317 p=0.001	---
T3y	F=180.59 p=0	F=69.938 p=0	---
T4x	F=70.871 p=0	F=1.771 p=0.202	F=33.763 p=0
T4y	F=185.898 p=0	F=24.808 p=0	F=108.22 p=0

Tableau 6.29 Résultats d'analyses ANOVA pour les sujets français

	PB			CV			AV		
/t/	/itu/ vs /ita/	/iti/ vs /ita/	/itu/ vs /iti/	/itu/ vs /ita/	/iti/ vs /ita/	/itu/ vs /iti/	/itu/ vs /ita/	/iti/ vs /ita/	/itu/ vs /iti/
T1x		0,08		---	---	---		0,05	
T1y				---	---	---			
T2x	0,10		0,07	0,27	-0,06	0,34		-0,05	
T2y		0,34	-0,32	0,10	0,21	-0,11		0,49	-0,49
T3x	0,08	0,15	-0,07	0,16	-0,09	0,25	---	---	---
T3y	0,13	0,68	-0,55	0,18	0,56	-0,38	---	---	---
T4x	0,17		0,16	0,06		0,09	0,08	-0,19	0,27
T4y	0,44	0,40		0,45	0,42		0,52	-0,09	0,61

Tableau 6.30 Résultats de tests post-hoc pour les sujets français

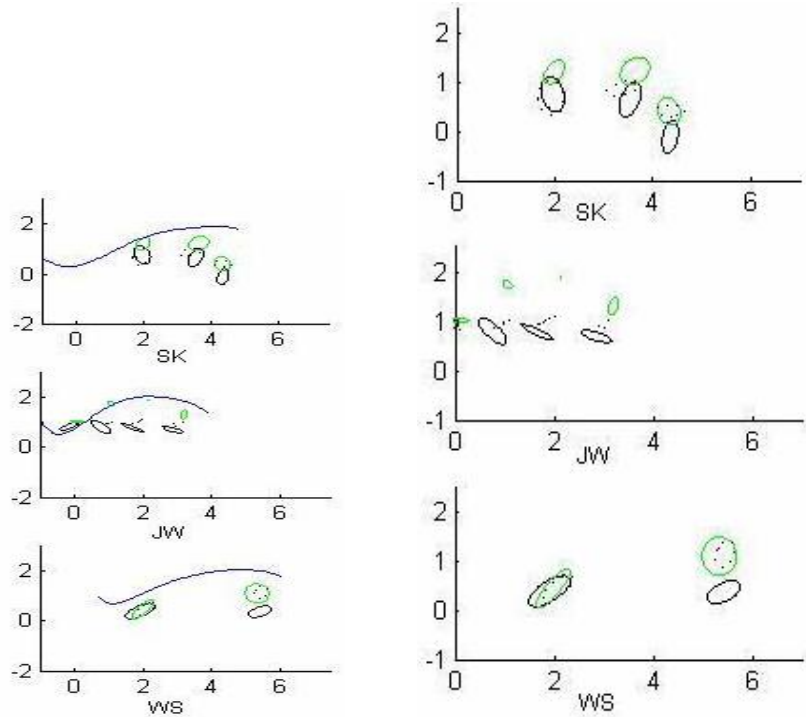


Figure 6.17 : Positions des capteurs (en cm) pour la consonne C = /t/ dans les séquences /ita/, /iti/ et /itu/ pour les trois sujets chinois

/t / in /ita/, /iti/, /itu/	SK	JW	WS
T1x	---	F=2.948 p=0.163	F=16.402 p=0
T1y	---	F=7.973 p=0.040	F=5.096 p=0.018
T2x	F=0.776 p=0.477	F=7.289 p=0.046	---
T2y	F=22.725 p=0	F=80.651 p=0.001	---
T3x	F=10.841 p=0.001	F=9.174 p=0.032	---
T3y	F=34.104 p=0	F=247.81 p=0	---
T4x	F=0.399 p=0.678	F=5.467 p=0.042	F=5.632 p=0.013
T4y	F=39.325 p=0	F=52.741 p=0.001	F=188.71 p=0

Tableau 6.31 Résultats d'analyses ANOVA pour les sujets chinois

	SK			JW			WS		
/t/	/itu/ vs /ita/	/iti/ vs /ita/	/itu/ vs /iti/	/itu/ vs /ita/	/iti/ vs /ita/	/itu/ vs /iti/	/itu/ vs /ita/	/iti/ vs /ita/	/itu/ vs /iti/
T1x	---	---	---	0,23	0,25		0,19	0,08	0,11
T1y	---	---	---	0,10	0,18	-0,08	0,08	0,06	
T2x				0,21	0,30	-0,09	---	---	---
T2y	-0,06	0,45	-0,51	0,16	0,95	-0,79	---	---	---
T3x	-0,17	0,10	-0,27	0,17	0,47	-0,30	---	---	---
T3y	0,22	0,58	-0,36	0,24	1,09	-0,86	---	---	---
T4x				0,16	0,35	-0,19		-0,09	0,12
T4y	0,54	0,52		0,23	0,61	-0,39	0,78	0,72	0,05

Tableau 6.32 Résultats de tests post-hoc pour les sujets chinois

6.2.3 Position atteinte par la langue pour C=/t/ dans les séquences /uta/, /uti/ et /utu/.

a. Sujets français

La figure 6.18 représente les positions de la langue pour C = /t/ dans les séquences /uta/, /uti/ et /utu/ pour les trois sujets français. Les tableaux 6.33 et 6.34 montrent les résultats d'analyses ANOVA et les résultats de tests post-hoc.

Observations :

- Nous retrouvons la même influence de V_2 en contexte $V_1=/u/$ qu'en contexte $V_1=/a/$ (figure 6.14). Les différences principales associées à V_2 portent sur la position verticale pour T_2 et/ou pour T_3 chez les trois sujets. Cette position est plus haute pour $V_2=/i/$, que pour $V_2 = /u /$ et pour $V_2=/a/$. La différence moyenne pour T_{2y} : (/uti/-/uta/) est 3.9mm pour PB ,4.3mm pour CV et 6.5mm pour AV; (/uti/-/utu/) est 4.6mm pour PB ,6.0mm pour CV et 5.3mm pour AV. pour T_{3y} : (/uti/-/uta/) est 5.2mm pour PB et 8.4mm pour CV; (/uti/-/utu/) est 4.5mm pour PB et 9.0mm pour CV.

b. Sujets chinois

La figure 6.19 représente les positions de la langue pour C = /t/ dans les séquences /uta/, /uti/ et /utu/ pour les trois sujets chinois. Les tableaux 6.35 et 6.36 montrent les résultats d'analyses ANOVA et les résultats de tests post-hoc.

Observations :

- La grande influence de V_2 sur /t/ est retrouvée en contexte $V_1=/u/$. Les différences associées à V_2 sont très remarquables sur la position verticale de T_2 , de T_3 et de T_4 chez trois sujets. Cette position est plus haute pour $V_2=/i /$, que pour $V_2=/u/$ et pour $V_2=/a/$. La différence moyenne pour T_{2y} : (/uti/-/uita/) est 9.3mm pour SK et 8.6mm pour JW; (/uti/-/utu/) est 8.2mm pour SK et 8.4mm pour JW. La différence moyenne pour T_{3y} : (/uti/-/uita/) est 8.6mm pour SK et 11.7mm pour JW; (/uti/-/utu/) est 5.5mm pour SK et 10.1mm pour JW. pour T_{4y} : (/uti/-/uita/) est 7.9mm pour SK, 6.7mm pour JW et 6.7mm pour WS; (/uti/-/utu/) est 2.7mm pour SK, 4.4mm pour JW et 1.2mm pour WS. (/utu/-/uta/) est 5.2mm pour SK, 2.3mm pour JW et 5.5mm pour WS.

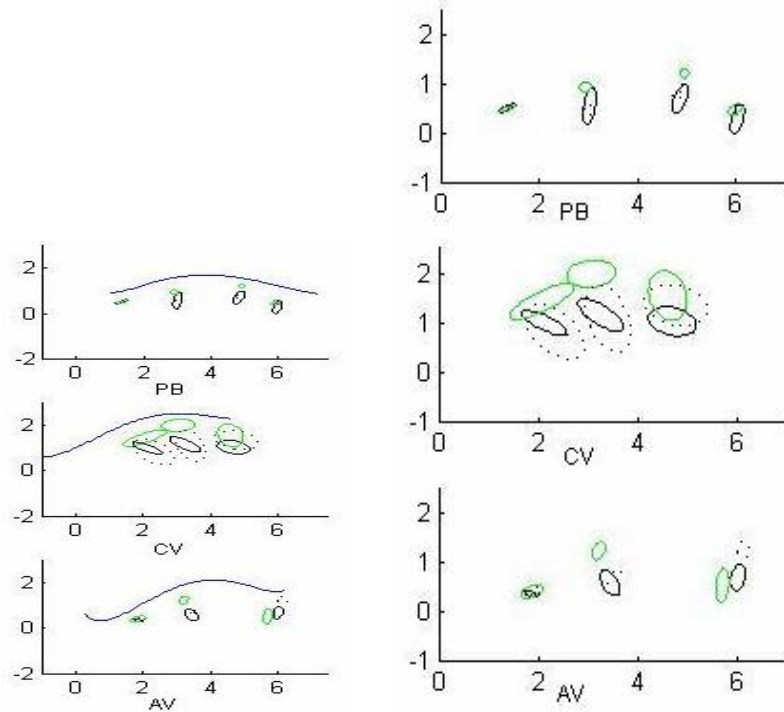


Figure 6.18 : Positions des capteurs (en cm) pour la consonne C = /t/ dans les séquences /uta/, /uti/ et /utu/ pour les trois sujets français

/t/ in /uta/, /uti/, /utu/	PB	CV	AV
T1x	F=0.621 p=0.548	---	F=4.580 p=0.027
T1y	F=1.226 p=0.317	---	F=3.487 p=0.055
T2x	F=7.485 p=0.004	F=4.687 p=0.025	F=61.901 p=0
T2y	F=45.487 p=0	F=17.139 p=0	F=167.23 p=0
T3x	F=23.665 p=0	F=8.291 p=0.003	---
T3y	F=108.97 p=0	F=38.366 p=0	---
T4x	F=2.158 p=0.145	F=1.182 p=0.332	F=117.98 p=0
T4y	F=16.651 p=0	F=40.71 p=0	F=155.52 p=0

Tableau 6.33 Résultats d'analyses ANOVA pour les sujets français

	PB			CV			AV		
/t/	/utu/ vs /uta/	/uti/ vs /uta/	/utu/ vs /uti/	/utu/ vs /uta/	/uti/ vs /uta/	/utu/ vs /uti/	/utu/ vs /uta/	/uti/ vs /uta/	/utu/ vs /uti/
T1x				---	---	---	0,11		0,09
T1y				---	---	---			
T2x		-0,08		0,21	-0,05	0,27	0,07	-0,23	0,30
T2y	-0,07	0,39	-0,46	-0,17	0,43	-0,60	0,13	0,65	-0,53
T3x	-0,07	0,09	-0,16	0,16	-0,19	0,35	---	---	---
T3y	0,07	0,52	-0,45	-0,06	0,84	-0,90	---	---	---
T4x				0,08	-0,09	0,17	0,11	-0,31	0,42
T4y	0,22	0,20		0,32	0,53	-0,20	0,57	-0,16	0,72

Tableau 6.34 Résultats de tests post-hoc pour les sujets français

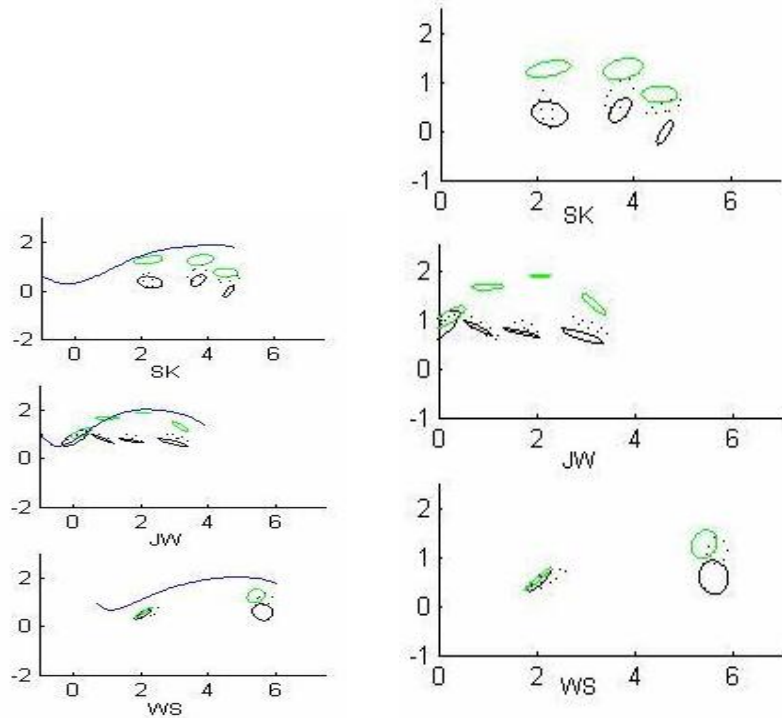


Figure 6.19 : Positions des capteurs (en cm) pour la consonne C = /t/ dans les séquences /uta/, /uti/ et /utu/ pour les trois sujets chinois

/t/ in /uta/,/uti/,/utu/	SK	JW	WS
T1x	---	F=4.556 p=0.093	F=16.966 p=0
T1y	---	F=7.774 p=0.042	F=1.645 p=0.224
T2x	F=0.365 p=0.705	F=10.161 p=0.027	---
T2y	F=116.25 p=0	F=161.13 p=0	---
T3x	F=1.069 p=0.388	F=14.173 p=0.015	---
T3y	F=50.373 p=0	F=956.04 p=0	---
T4x	F=0.173 p=0.844	F=4.266 p=0.102	F=54.303 p=0
T4y	F=79.321 p=0	F=1009.2 p=0	F=84.705 p=0

Tableau 6.35 Résultats d'analyses ANOVA pour les sujets chinois

	SK			JW			WS		
/t/	/utu/ vs /uta/	/uti/ vs /uta/	/utu/ vs /uti/	/utu/ vs /uta/	/uti/ vs /uta/	/utu/ vs /uti/	/utu/ vs /uta/	/uti/ vs /uta/	/utu/ vs /uti/
T1x	---	---	---		0,17		0,19		0,19
T1y	---	---	---	0,14	0,21	-0,08			
T2x				0,07	0,19	-0,12	---	---	---
T2y	0,10	0,93	-0,82		0,86	-0,84	---	---	---
T3x				0,06	0,35	-0,29	---	---	---
T3y	0,30	0,86	-0,55	0,15	1,17	-1,01	---	---	---
T4x					0,20		0,08	-0,19	0,27
T4y	0,52	0,79	-0,27	0,23	0,67	-0,44	0,55	0,67	-0,12

Tableau 6.36 Résultats de tests post-hoc pour les sujets chinois

Résumé :

- Ü Pour les six sujets français et chinois, nous avons observé une influence significative et très remarquable associée à V_2 sur la position verticale de T_2 et T_3 de la consonne /k/. L'amplitude de cette influence liée aux changements de V_2 est aussi grande pour les sujets chinois que pour les sujets français. Elle est cohérente avec l'hypothèse de l'anticipation de la voyelle suivante sur la consonne.
- Ü En ce qui concerne la position verticale du capteur T_4 , les trois sujets français montrent une position plus haute pour $V_2=/u/$ que pour $V_2=/a/$ et pour $V_2=/i/$. Ceci est cohérent avec le fait que la position du corps de langue est plus haute pour une voyelle /u/. Mais pour les trois sujets chinois, la position verticale du T_4 est plus haute pour $V_2=/i/$ que pour $V_2=/a/$ et pour $V_2=/u/$ dans toutes les séquences. Cela est en contradiction avec une stratégie d'anticipation.
- Ü Pour les six sujets, l'influence associée à V_2 sur la direction horizontale est également observée, mais elle n'est pas toujours significative. Certaines influences significatives sont en contradiction avec une anticipation. Nous pensons que cela vient des contraintes associées à la production de la consonne occlusive alvéolaire, qui exige une occlusion dans la partie très antérieure du conduit vocal.

6.2.4 Position atteinte par la langue pour $C=/k/$ dans les séquences /aka/ et /aku/

a. Sujets français

La figure 6.20 représente les positions de la langue pour $C = /k/$ dans les séquences /aka/ et /aku/ pour les trois sujets français. Les tableaux 6.37 et 6.38 montrent les résultats d'analyses ANOVA et les résultats de tests post-hoc.

Observations :

- Les différences associées à V_2 sur la position horizontale pour T_1 , T_2 , T_3 et T_4 sont observées pour tous les trois sujets. La position de la langue est plus en arrière pour $V_2=/u/$, que pour $V_2=/a/$. pour T_{1x} : (/aku/-/aka/) est 3.6mm pour PB et 3.8mm pour AV; pour T_{2x} : (/aku/-/aka/) est 3.5mm pour PB ,7.0mm pour CV et 6.0mm pour AV;

La différence moyenne pour T_{3x} : (/aku/-/aka/) est 2.6mm pour PB et 7.5mm pour CV; de T_{4x} : (/aku/-/aka/) est 2.8mm pour PB, 7.6mm pour CV et 2.5mm pour AV.

- Les sujets PB et AV montrent des différences sur la position verticale de T_1 , T_2 , T_3 et T_4 , mais elles sont moins claires que sur la position horizontale. En contradiction avec une anticipation, la position verticale de la langue est plus haute pour $V_2=/a/$, que pour $V_2 = /u/$. La différence moyenne pour T_{1y} : (/aka/-/aku/) est 3.0mm pour PB et 3.7mm pour AV; pour T_{2y} : (/aka/-/aku/) est 3.4mm pour PB et 3.5mm pour AV; Pour T_{3y} : (/aka/-/aku/) est 3.2mm pour PB; de T_{4y} : (/aka/-/aku/) est 1.3mm pour PB et 1.3mm pour AV.

b. Sujets chinois

La figure 6.21 représente les positions de la langue pour $C = /k/$ dans les séquences /aka/ et /aku/ pour les trois sujets chinois. Les tableaux 6.39 et 6.40 montrent les résultats d'analyses ANOVA et les résultats de tests post-hoc.

Observations :

- Les différences sur la position horizontale sont parfois observées chez les sujets SK, JW et WS. La position de la langue est plus en arrière pour $V_2=/u/$, que pour $V_2=/a/$. La différence moyenne pour T_{1x} : (/aku/-/aka/) est 2.7mm pour le sujet WS; pour T_{2x} : (/aku/-/aka/) est 2.0mm pour le sujet SK ,2.1mm pour le sujet JW; La différence moyenne pour T_{3x} : (/aku/-/aka/) est 1.8mm pour le sujet JW; de T_{4x} : (/aku/-/aka/) est 1.4mm pour le sujet WS, 1.8mm pour le sujet JW.
- Les différences sur la position verticale sont aussi parfois observées chez les trois sujets. Nous retrouvons que la position verticale de la langue est plus haute pour $V_2=/a/$, que pour $V_2=/u/$. La différence moyenne pour T_{1y} : (/aka/-/aku/) est 2.0mm pour le sujet JW; de T_{2y} : (/aka/-/aku/) est 1.8 mm pour le sujet SK; pour T_{4y} : (/aka/-/aku/) est 1.4mm pour le sujet JW.

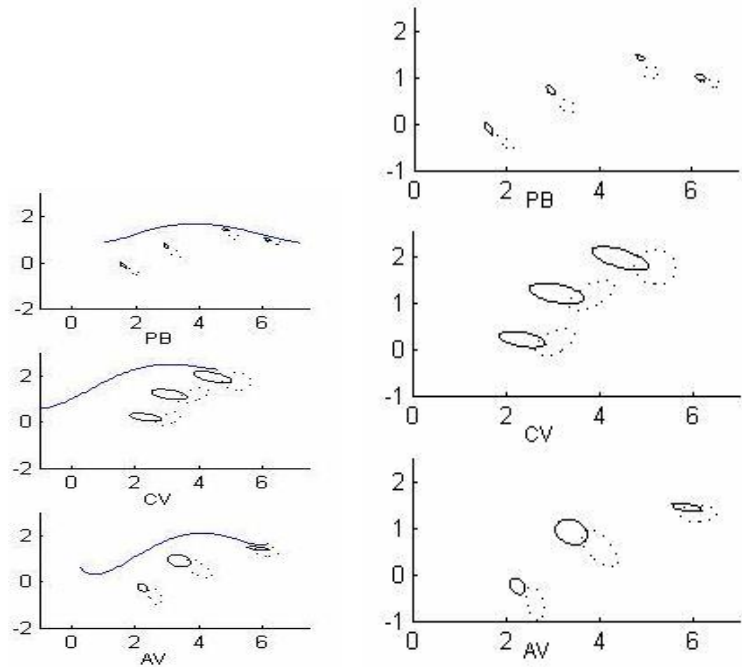


Figure 6.20 : Positions des capteurs (en cm) pour la consonne C = /k/ dans les séquences /aka/ et /aku/ pour les trois sujets français

/ k / in /aka/ /aku/	PB	CV	AV
T1x	F=144.10 p=0	---	F=123.93 p=0
T1y	F=81.272 p=0	---	F=30.292 p=0
T2x	F=544.72 p=0	F=50.93 p=0	F=97.741 p=0
T2y	F=160.06 p=0	F=1.873 p=0.213	F=85.54 p=0
T3x	F=75.089 p=0	F=52.204 p=0	---
T3y	F=193.6 p=0	F=0.980 p=0.355	---
T4x	F=108.29 p=0	F=58.925 p=0	F=13.45 p=0.005
T4y	F=75.278 p=0	F=19.198 p=0.003	F=26.07 p=0.001

Tableau 6.37 Résultats d'analyses ANOVA pour les sujets français

	PB	CV	AV
/k/ in /akV/	/aku/ vs /aka/	/aku/ vs /aka/	/aku/ vs /aka/
T1x	0,36	---	0,38
T1y	-0,30	---	-0,37
T2x	0,35	0,70	0,60
T2y	-0,34		-0,35
T3x	0,26	0,75	---
T3y	-0,32		---
T4x	0,28	0,76	0,25
T4y	-0,13	-0,19	-0,13

Tableau 6.38 Résultats de tests post-hoc pour les sujets français

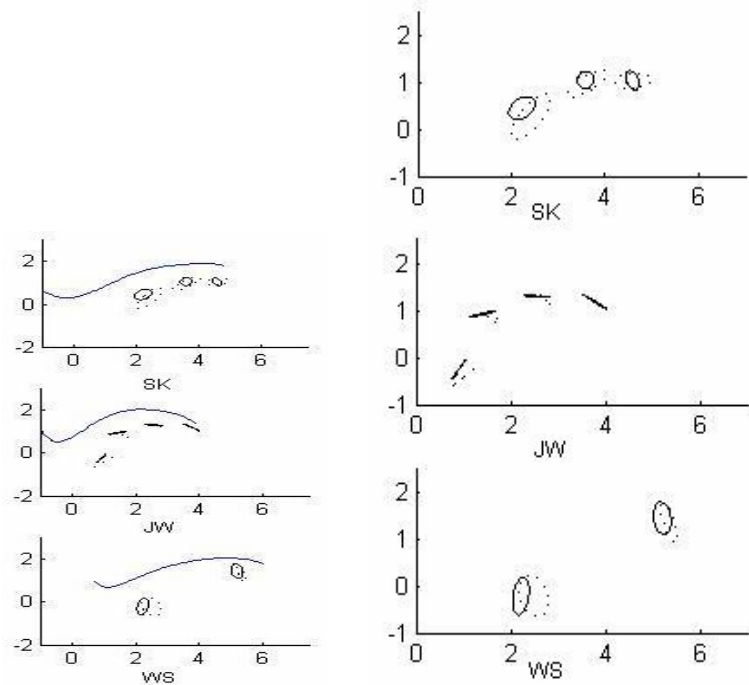


Figure 6.21 : Positions des capteurs (en cm) pour la consonne C = /k/ dans les séquences /aka/ et /aku/ pour les trois sujets chinois

/ k / in /aka/ /aku/	SK	JW	WS
T1x	---	F=0.216 p=0.688	F=20.719 p=0.002
T1y	---	F=6.280 p=0.036	F=0.071 p=0.797
T2x	F=9.552 p=0.018	F=27.00 p=0.035	---
T2y	F=6.481 p=0.047	F=0.670 p=0.499	---
T3x	F=0.292 p=0.606	F=13.21 p=0.048	---
T3y	F=1.806 p=0.221	F=3.610 p=0.196	---
T4x	F=0.057 p=0.817	F=17.587 p=0.050	F=10.508 p=0.012
T4y	F=0.002 p=0.969	F=25.090 p=0.038	F=1.743 p=0.223

Tableau 6.39 Résultats d'analyses ANOVA pour les sujets chinois

	SK	JW	WS
/k/ in /akV/ /aka/	/aku/ vs /aka/	/aku/ vs /aka/	/aku/ vs /aka/
T1x	---		0,27
T1y	---	-0,20	
T2x	0,20	0,21	---
T2y	-0,18		---
T3x		0,18	---
T3y			---
T4x		0,18	0.14
T4y		-0,14	

Tableau 6.40 Résultats de tests post-hoc pour les sujets chinois

6.2.5 Position atteinte par la langue pour C=/k/ dans les séquences /ika/ et /iku/

a. Sujets français

La figure 6.22 représente les positions de la langue pour C = /k/ dans les séquences /ika/ et /iku/ pour les trois sujets français. Les tableaux 6.41 et 6.42 montrent les résultats d'analyses ANOVA et les résultats de tests post-hoc.

Observations :

- Nous avons trouvé le même effet chez les sujets français. Des différences associées à V₂ sur la position horizontale de T₁, T₂, T₃ et T₄ sont clairement observées. La position horizontale est plus en arrière pour V₂=/u/, que pour V₂=/a/. La différence moyenne pour T_{1x} : (/iku/-/ika/) est 3.5mm pour le sujet PB et 3.3mm pour le sujet AV; pour T_{2x} : (/iku/-/ika/) est 3.7mm pour le sujet PB ,6.1mm pour le sujet CV et 4.5mm pour le sujet AV; La différence moyenne pour T_{3x} : (/iku/-/ika/) est 3.7mm pour le sujet PB et 6.0mm pour le sujet CV; pour T_{4x} : (/iku/-/ika/) est 3.9mm pour PB, 4.7mm pour CV et 2.9mm pour le sujet AV.
- Pour les trois sujets, nous notons que la position verticale de la langue est plus haute pour V₂=/a/, que pour V₂=/u/. La différence moyenne pour T_{1y} : (/ika/-/iku/) est 1.9mm pour PB et 2.8mm pour AV; pour T_{2y} : (/iku/-/ika/) est 1.9mm pour PB et 4.5mm pour AV ; Pour T_{3y} : (/iku/-/ika/) est 1.9mm pour PB et 1.8mm pour CV; pour T_{4y} : (/iku/-/ika/) est 1.6mm pour PB et 2.0mm pour CV.

b. Sujets chinois

La figure 6.23 représente les positions de la langue pour C = /k/ dans les séquences /ika/ et /iku/ pour les trois sujets chinois. Les tableaux 6.43 et 6.44 montrent les résultats d'analyses ANOVA et les résultats de tests post-hoc.

Observations :

- Des différences associées à V₂ sur la position horizontale sont parfois observées chez SK et WS. La position de la langue est plus en arrière pour V₂=/u/, que pour V₂ = /a/. La différence moyenne pour T_{1x} : (/iku/-/ika/) est 3.6mm pour WS; pour T_{2x} : (/iku/-/ika/) est 2.6mm pour SK;
- Le sujet JW ne montre pas d'effet.

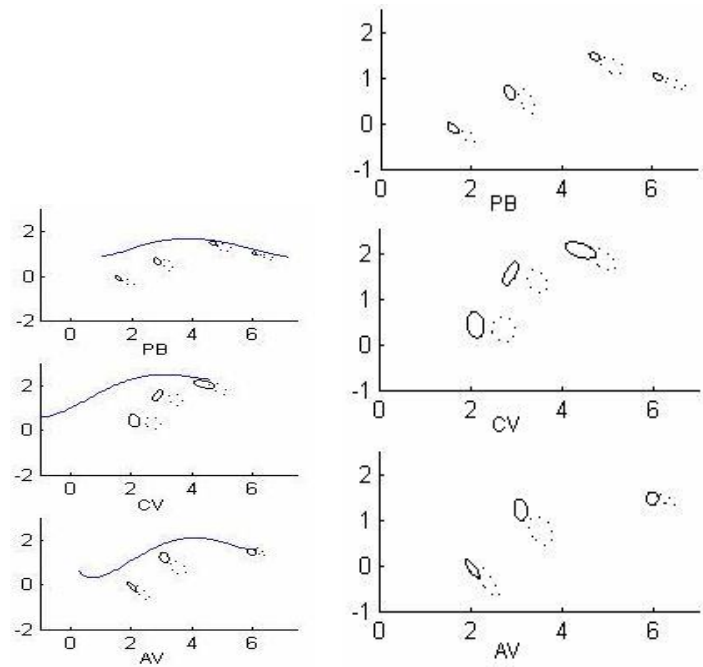


Figure 6.22 : Positions des capteurs (en cm) pour la consonne C = /k/ dans les séquences /ika/ et /iku/ pour les trois sujets français

/ k / in /ika/ /iku/	PB	CV	AV
T1x	F=171.53 p=0	---	F=59.919 p=0
T1y	F=133.87 p=0	---	F=17.241 p=0.002
T2x	F=104.51 p=0	F=221.05 p=0	F=69.712 p=0
T2y	F=21.707 p=0	F=3.108 p=0.116	F=56.1 p=0
T3x	F=50.158 p=0	F=185.74 p=0	---
T3y	F=61.153 p=0	F=16.354 p=0.004	---
T4x	F=71.371 p=0	F=11.29 p=0	F=30.193 p=0
T4y	F=72.437 p=0	F=78.767 p=0	F=2.216 p=0.171

Tableau 6.41 Résultats d'analyses ANOVA pour les sujets français

	PB	CV	AV
/ k / in /ikV/	/iku/ vs /ika/	/iku/ vs /ika/	/iku/ vs /ika/
T1x	0,35	---	0,33
T1y	-0,19	---	-0,28
T2x	0,37	0,61	0,45
T2y	-0,19		-0,45
T3x	0,37	0,60	---
T3y	-0,19	-0,18	---
T4x	0,39	0,47	0,29
T4y	-0,16	-0,20	

Tableau 6.42 Résultats de tests post-hoc pour les sujets français

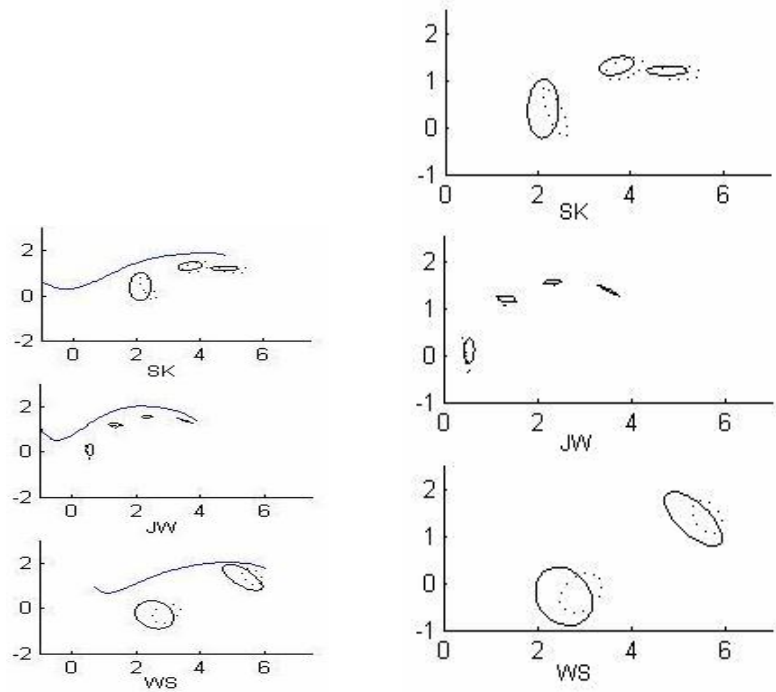


Figure 6.23 : Positions des capteurs (en cm) pour la consonne C = /k/ dans les séquences /ika/ et /iku/ pour les trois sujets chinois

/ k / in /ika/ /iku/	SK	JW	WS
T1x	---	F=2.674 p=0.244	F=28.84 p=0
T1y	---	F=6.250 p=0.130	F=0.845 p=0.382
T2x	F=34.979 P=0	F=0.086 p=0.797	---
T2y	F=0.347 P=0.572	F=12.64 p=0.071	---
T3x	F=5.358 P=0.044	F=0.082 p=0.802	---
T3y	F=0.822 P=0.391	F=0.250 p=0.130	---
T4x	F=7.980 P=0.022	F=0.018 p=0.904	F=17.805 p=0.002
T4y	F=3.368 P=0.104	F=0.158 p=0.729	F=0.180 p=0.681

Tableau 6.43 Résultats d'analyses ANOVA pour les sujets chinois

	SK	JW	WS
/ k / in /ikV/	/iku/ vs /ika/	/iku/ vs /ika/	/iku/ vs /ika/
T1x	---		0.36
T1y	---		
T2x	0,26		---
T2y			---
T3x	0,19		---
T3y			---
T4x	0,26		0.30
T4y			

Tableau 6.44 Résultats de tests post-hoc pour les sujets chinois

6.2.6 Position atteinte par la langue pour C=/k/ dans les séquences /uka/ et /uku/

a. Sujets français

La figure 6.24 représente les positions de la langue pour C = /k/ dans les séquences /uka/ et /uku/ pour les trois sujets français. Les tableaux 6.45 et 6.46 montrent les résultats d'analyses ANOVA et les résultats de tests post-hoc.

Observations :

- Des différences sur la position horizontale de T₁, T₂, T₃ et T₄ sont observées chez les trois sujets. La position de la langue est plus en arrière pour V₂=/u/, que pour V₂=/a/. La différence moyenne pour T_{1x} : (/uku/-/uka/) est 3.3mm pour PB; pour T_{2x} : (/uku/-/uka/) est 3.9mm pour PB, 3.2mm pour CV et 2.8mm pour AV; pour T_{3x} : (/uku/-/uka/) est 4.0mm pour PB et 3.2mm pour CV; pour T_{4x} : (/uku/-/uka/) est 4.8mm pour PB, 2.5mm pour CV.
- La position verticale de la langue est plus haute pour V₂=/a/, que pour V₂=/u/. La différence moyenne pour T_{1y} : (/uka/-/uku/) est 2.5mm pour PB et 1.0mm pour AV; pour T_{2y} : (/uku/-/uka/) est 2.6mm pour PB et 3.3mm pour AV; pour T_{3y} : (/uku/-/uka/) est 2.5mm pour PB et 1.8mm pour CV; pour T_{4y} : (/uku/-/uka/) est 1.5mm pour PB et 2.2mm pour CV.

b. Sujets chinois

La figure 6.25 représente les positions de la langue pour C = /k/ dans les séquences /uka/ et /uku/ pour les trois sujets chinois. Les tableaux 6.47 et 6.48 montrent les résultats d'analyses ANOVA et les résultats de tests post-hoc.

Observations :

- La position horizontale de la langue est plus en arrière pour V₂=/u/, que pour V₂=/a/. La différence moyenne pour T_{1x} : (/uku/-/uka/) est 3.5mm pour WS; pour T_{2x} : (/uku/-/uka/) est 1.2mm pour JW ; pour T_{4x} : (/uku/-/uka/) est 3.4mm pour WS.
- La position verticale de la langue est plus haute pour V₂=/a/, que pour V₂ = /u/. La différence moyenne pour T_{2y} : (/uku/-/uka/) est 2.7mm pour SK et 1.0mm pour JW; pour T_{3y} : (/uku/-/uka/) est 2.0mm pour SK et 1.0mm pour JW; pour T_{4y} : (/uku/-/uka/) est 1.6mm pour SK et 1.4mm pour WS.

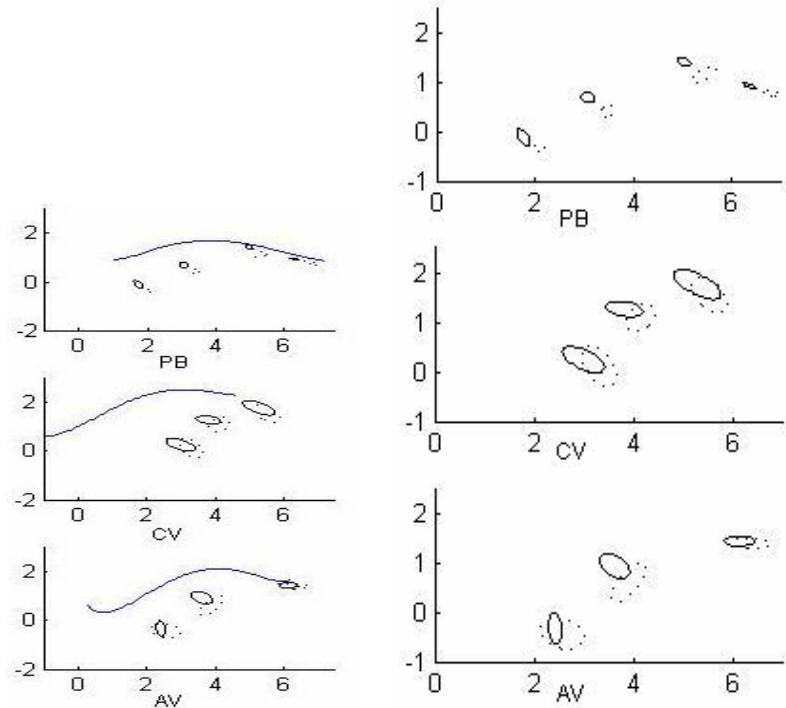


Figure 6.24 : Positions des capteurs (en cm) pour la consonne C = /k/ dans les séquences /uka/ et /uku/ pour les trois sujets français

/ k / in /uka/,/uku/	PB	CV	AV
T1x	F=162.53 P=0	---	F=3.675 p=0.092
T1y	F=131.29 P=0	---	F=6.444 p=0.035
T2x	F=345.15 P=0	F=8.802 p=0.021	F=9.232 p=0.016
T2y	F=94.20 P=0	F=1.746 p=0.228	F=56.817 p=0
T3x	F=72.18 P=0	F=14.513 p=0.007	---
T3y	F=75.763 P=0	F=10.239 p=0.015	---
T4x	F=163.08 P=0	F=5.596 p=0.050	F=5.142 p=0.053
T4y	F=148.89 P=0	F=15.937 p=0.005	F=2.006 p=0.194

Tableau 6.45 Résultats d'analyses ANOVA pour les sujets français

	PB	CV	AV
/ k / in /ukV/	/uku/ vs /uka/	/uku/ vs /uka/	/uku/ vs /uka/
T1x	0,33	---	
T1y	-0,25	---	-0,10
T2x	0,39	0,32	0,28
T2y	-0,26		-0,33
T3x	0,40	0,32	---
T3y	-0,25	-0,18	---
T4x	0,48	0,25	
T4y	-0,15	-0,22	

Tableau 6.46 Résultats de tests post-hoc pour les sujets français

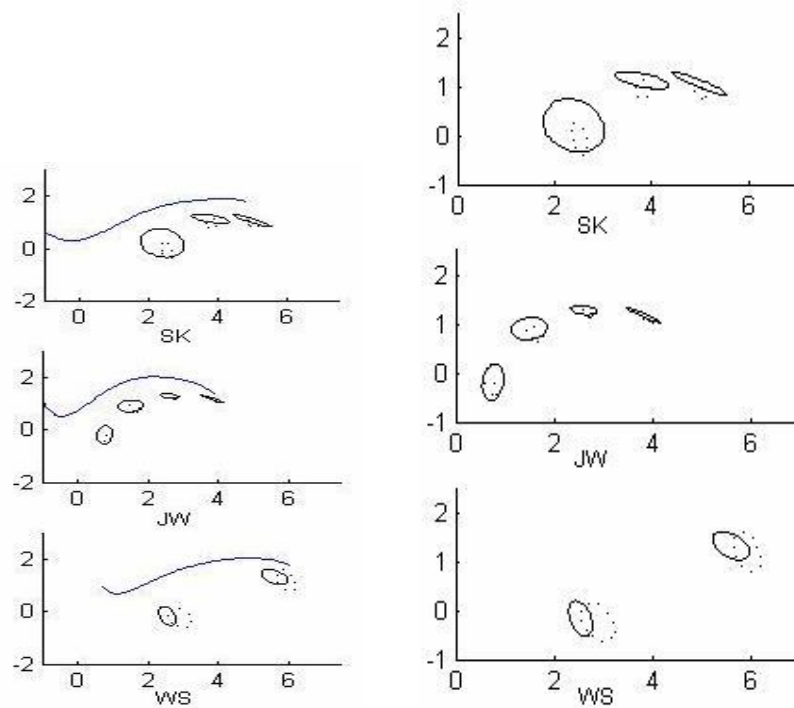


Figure 6.25 : Positions des capteurs (en cm) pour la consonne C = /k/ dans les séquences /uka/ et /uku/ pour les trois sujets chinois

/ k / in /uka/,/uku/	SK	JW	WS
T1x	---	F=0.007 $p=0.941$	F=36.284 $p=0$
T1y	---	F=3.008 $p=0.225$	F=5.909 $p=0.038$
T2x	F=1.253 $p=0.314$	F=12.016 $p=0.049$	---
T2y	F=5.814 $p=0.050$	F=15.211 $p=0.050$	---
T3x	F=0.044 $p=0.842$	F=2.191 $p=0.277$	---
T3y	F=36.858 $p=0.002$	F=75.00 $p=0.013$	---
T4x	F=0.025 $p=0.880$	F=2.729 $p=0.240$	F=31.202 $p=0$
T4y	F=31.557 $p=0.002$	F=12.964 $p=0.069$	F=23.963 $p=0.001$

Tableau 6.47 Résultat d'analyses ANOVA pour les sujets chinois

	SK	JW	WS
/ k / in /ukV/	/uku/ vs /uka/	/uku/ vs /uka/	/uku/ vs /uka/
T1x	---		0,35
T1y	---		-0,07
T2x		0,12	---
T2y	-0,27	-0,10	---
T3x			---
T3y	-0,20	-0,10	---
T4x			0,34
T4y	-0,16		-0,14

Tableau 6.48 Résultat de tests post-hoc pour les sujets chinois

Résumé :

- Ü Des variabilités significatives associées à V_2 sont clairement observées dans tous les cas pour les trois sujets français selon la direction horizontale. La position de la langue est plus en arrière pour $V_2=/u/$ que pour $V_2=/a/$. Cela est cohérent avec une anticipation de la voyelle dans la consonne précédente.
- Ü Pour les sujets chinois, des variabilités selon la direction horizontale sont observées en fonction de V_2 pour la consonne $/k/$, mais elles sont moins grandes et moins systématiques d'un sujet à l'autre que pour la consonne $/t/$ (cf.section précédente).
- Ü Des variabilités selon la direction verticale sont aussi observées pour les six sujets, mais elles sont dans la plupart des cas en contradiction avec une stratégie d'anticipation. Nous pensons que cela est dû à la combinaison d'influences de différentes stratégies d'anticipation du locuteur, contraintes perceptives et contraintes physiques pour la consonne vélaire $/k/$. Nous reviendrons sur ce point dans le chapitre 8.

Bilan générale :

Nous avons observé une différence entre les locuteurs français et les locuteurs chinois dans les deux sections précédentes. Pour la coarticulation d'anticipation V-à-V, nos données montrent que les locuteurs français sont capables d'anticiper la production de la voyelle V_2 pendant la production de la voyelle V_1 , mais ce n'est pas le cas pour les locuteurs chinois. Pour la coarticulation d'anticipation V-à-C, les six locuteurs montrent que la consonne C est influencée par la voyelle V_2 . Les résultats montrent que, pour les locuteurs français, la voyelle V_2 influence la séquence entière de V_1CV_2 , tandis que cette influence est limitée dans la syllabe CV_2 pour les locuteurs chinois. Cela pourrait être dû aux différentes stratégies choisies par locuteurs français et par locuteurs chinois. Comme nous avons supposé dans le chapitre 2 que, le statut de la syllabe ne semble pas avoir la même importance dans la parole. Ces différentes stratégies pourraient être la conséquence respective de statut de la syllabe dans leur langue. Pour soutenir cette hypothèse, nous essayons d'exclure les autres possibilités qui peuvent affecter la différence que nous avons observée sur l'anticipation V-à-V entre les deux

groupes de langue (voir aussi une discussion générale dans le chapitre 8).

- (1) La densité de l'inventaire vocalique a été suggérée comme pouvant influencer la coarticulation de voyelle à voyelle (Manuel, 1990). Nous avons décrit l'inventaire vocalique en chinois et en français dans le chapitre 2. Le chinois a principalement 5 voyelles tandis que le français en a 11 dans son système vocalique. Selon l'hypothèse de Manuel, la variabilité de V1 en chinois devrait être plus grande que celle en français, Ce qui est contraire à nos observations. Donc cette hypothèse ne peut pas expliquer nos résultats.
- (2) L'anticipation diminue quand la durée de transition entre les cibles augmente (Abry and Lallouache 1996). Donc la durée de transition V_1-V_2 peut influencer la coarticulation de voyelle à voyelle. Nous menons une analyse temporelle dans la section suivante pour exclure cette possibilité.

6.3 Analyse temporelle

L'analyse temporelle a pour but de vérifier si la durée de transition V_1-V_2 influence l'anticipation V_1-V_2 que nous avons observée dans les données articulatoires. En chinois, la consonne occlusive peut être aspirée ou non aspirée. Dans notre expérience, nous constatons que les locuteurs chinois ont interprété les consonnes /t/ et /k/ comme consonnes aspirés /t^h/ et /k^h/. Par conséquent, la durée de la consonne est potentiellement plus longue pour ces sujets. Si la durée de la transition V_1-V_2 est significativement différente entre les deux groupes de sujets, l'influence de cette durée sur la différence d'anticipation observée chez locuteur français et chez locuteur chinois sera non négligeable.

En reprenant l'hypothèse d'Abry selon laquelle le temps pour atteindre la cible influence l'anticipation, cette analyse temporelle se concentre sur la durée entre les cibles de phonèmes. Nous reprenons ici la cible articulatoire atteinte. Nous parlons dans le chapitre 5 de tous les étiquetages que nous avons faits afin de trouver la cible articulatoire atteinte (c'est à dire le point le plus extrême des articulateurs). Afin d'analyser le temps mis à atteindre la cible V_2 , nous proposons deux durées de transitions entre les cibles d'une séquence V_1CV_2 : une durée de la transition V_1-V_2 et une durée de transition $C-V_2$.

La première durée caractérise l'intervalle de la voyelle V_1 à la voyelle V_2 . Elle est mesurée à partir de la cible articulatoire V_1 à la cible articulatoire V_2 . La deuxième durée caractérise le geste d'ouverture à partir de la consonne C à la voyelle V_2 . Elle est mesurée du

début de l'explosion consonantique C à la cible articulatoire V₂.

- Durées voyelle à voyelle : $\text{Durée_V}_1\text{V}_2 = X_{V_2} - X_{V_1}$;
- Durées du geste d'ouverture vocalique : $\text{Durée_CV}_2 = X_{V_2} - X_C$;

Des mesures temporelles peuvent être effectuées grâce aux étiquettes précédemment positionnées (voir : l'étiquetage dans le chapitre 5). Dans un premier temps, il s'agit de calculer les durées absolues en millisecondes de ces étiquetages. (Figure 6.26)

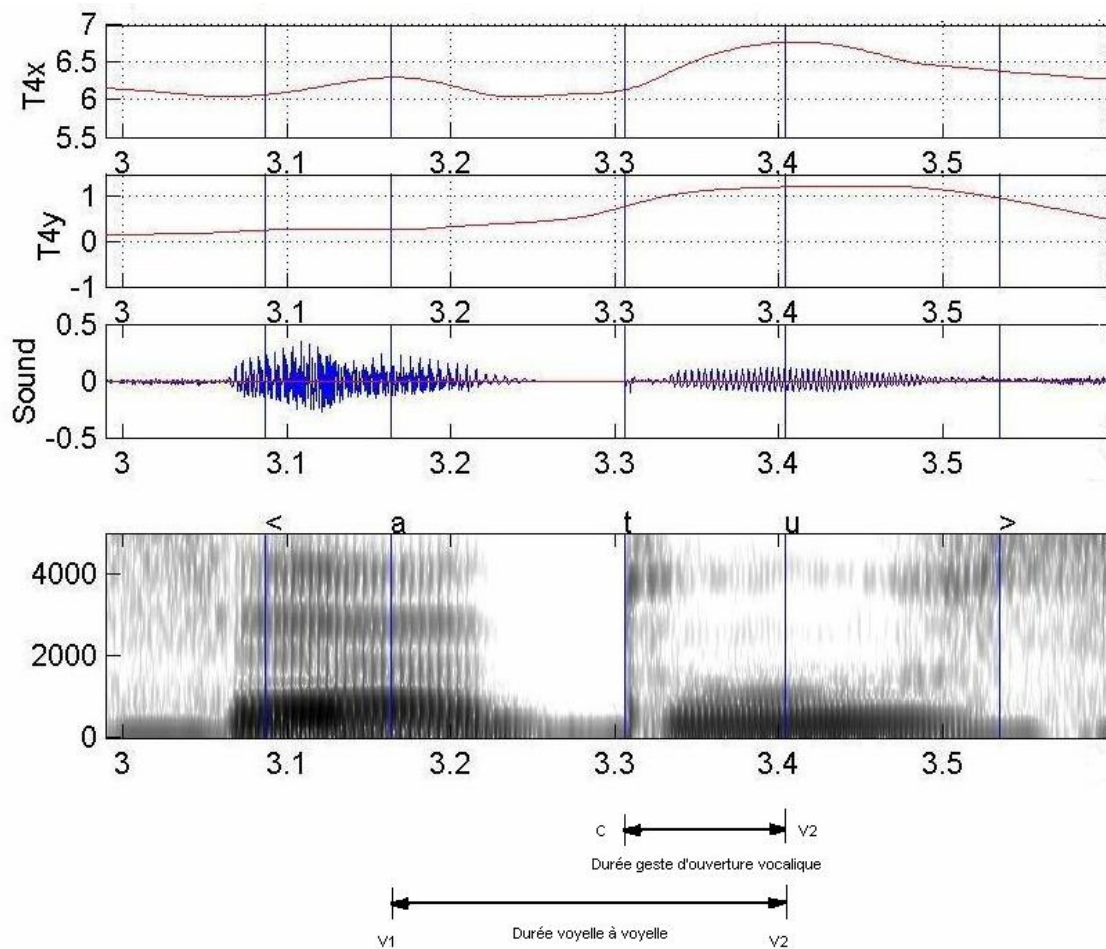


Figure 6.26 : Les paramètres de durée. (Voir figure 5.2 pour les détails) Durée CV₂ : durée du geste d'ouverture vocalique. Durée V₁V₂ : durée voyelle à voyelle.

Les analyses de variance de type ANOVA Univariée (inter-sujets) ont été effectuées. Les variables dépendantes étaient la Durée CV₂ et la Durée V₁V₂, et la variable indépendante était les six sujets. Toutes les analyses ont été faites sur la consonne vélaire /k/ et alvéolaire /t/ respectivement. Le logiciel SPSS pour Windows a été utilisé pour cette analyse. Les tests de post-hoc LSD, où toutes les moyennes ont été comparées deux à deux, ont été effectués après les analyses d'Univariée. Nous ne présenterons que les résultats significatifs ($p < 0.05$).

6.3.1 Analyse de la durée entre la consonne C et la cible voyelle V₂

6.3.1.1 Séquences /V₁tV₂/

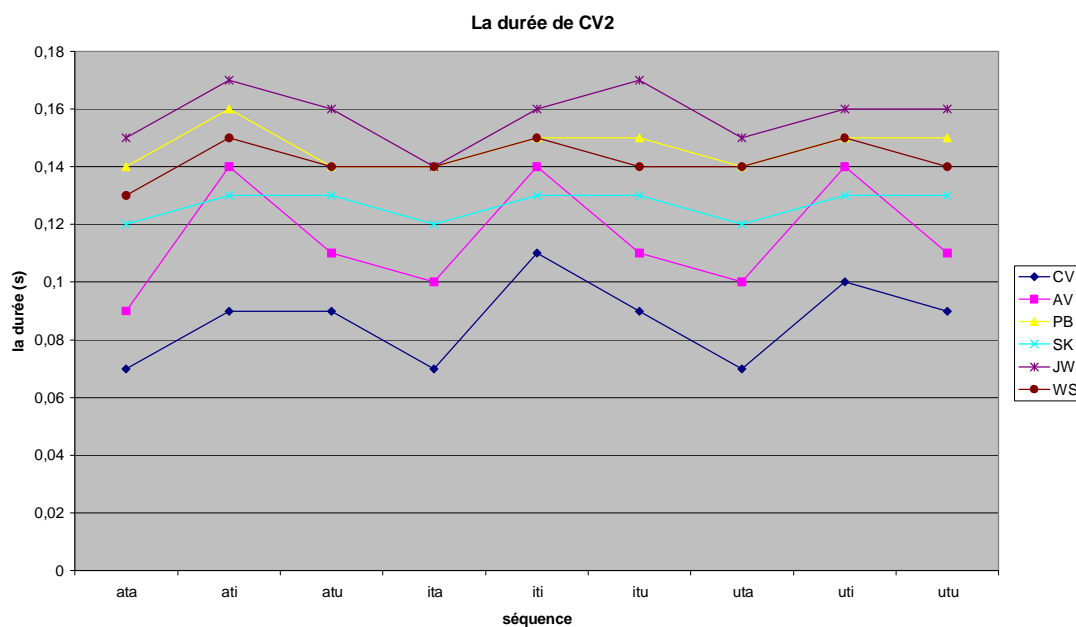


Figure 6.27 Durée moyenne entre la consonne C et la cible voyelle V₂ dans les séquences /V₁tV₂/ pour les six sujets.

	ata(s)	ati(s)	atu(s)	ita(s)	iti(s)	itu(s)	uta(s)	uti(s)	utu(s)
Sujet : PB meanDurée_CV ₂	0,14	0,16	0,14	0,14	0,15	0,15	0,14	0,15	0,15
Sujet : PB Std Durée_CV ₂	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
Sujet : CV meanDurée_CV ₂	0,07	0,09	0,09	0,07	0,11	0,09	0,07	0,10	0,09
Sujet : CV Std Durée_CV ₂	0,01	0,01	0,01	0,01	0,02	0,01	0,01	0,02	0,01
Sujet : AV meanDurée_CV ₂	0,09	0,14	0,11	0,10	0,14	0,11	0,10	0,14	0,11
Sujet : AV Std Durée_CV ₂	0,01	0,02	0,01	0,01	0,02	0,01	0,01	0,02	0,01
Sujet : SK meanDurée_CV ₂	0,12	0,13	0,13	0,12	0,13	0,13	0,12	0,13	0,13
Sujet : SK Std Durée_CV ₂	0,02	0,02	0,01	0,01	0,02	0,01	0,01	0,02	0,01
Sujet : JW meanDurée_CV ₂	0,15	0,17	0,16	0,14	0,16	0,17	0,15	0,16	0,16
Sujet : JW Std Durée_CV ₂	0,02	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
Sujet : WS meanDurée_CV ₂	0,13	0,15	0,14	0,14	0,15	0,14	0,14	0,15	0,14
Sujet : WS Std Durée_CV ₂	0,02	0,01	0,01	0,02	0,01	0,01	0,02	0,02	0,01

Tableau 6.49 Durée moyenne et écart-type entre la consonne C et la cible voyelle V₂ (Durée_CV₂) dans les neuf séquences pour les six sujets

6.3.1.2 Séquences /V₁kV₂/

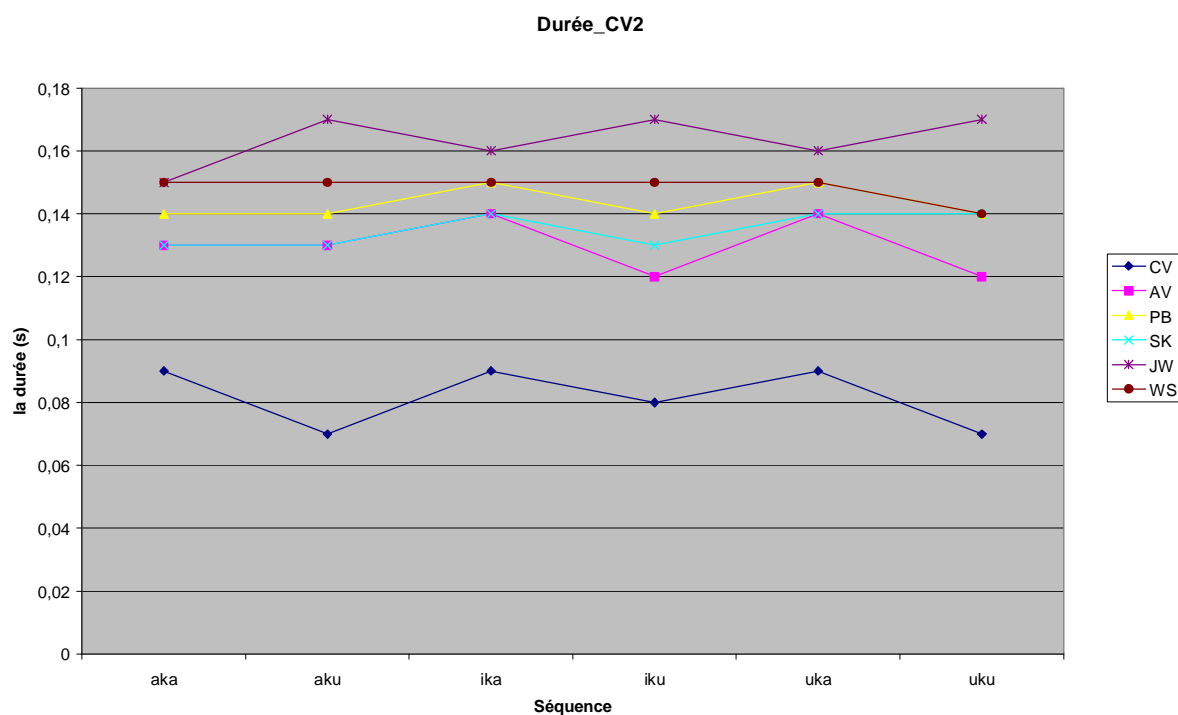


Figure 6.28 Durée moyenne entre la consonne C et la cible voyelle V₂ dans les séquences /V₁kV₂/ pour les six sujets.

	aka(s)	aku(s)	ika(s)	iku(s)	uka(s)	uku(s)
Sujet : PB						
meanDurée_CV ₂	0,14	0,14	0,15	0,14	0,15	0,14
Std Durée_CV ₂	0,01	0,01	0,01	0,01	0,01	0,01
Sujet : CV						
meanDurée_CV ₂	0,09	0,07	0,09	0,08	0,09	0,07
Std Durée_CV ₂	0,01	0,01	0,01	0,01	0,01	0,01
Sujet : AV						
meanDurée_CV ₂	0,13	0,13	0,14	0,12	0,14	0,12
Std Durée_CV ₂	0,02	0,01	0,01	0,02	0,01	0,02
Sujet : SK						
meanDurée_CV ₂	0,13	0,13	0,14	0,13	0,14	0,14
Std Durée_CV ₂	0,02	0,01	0,02	0,02	0,02	0,02
Sujet : JW						
meanDurée_CV ₂	0,15	0,17	0,16	0,17	0,16	0,17
Std Durée_CV ₂	0,02	0,01	0,01	0,01	0,01	0,01
Sujet : WS						
meanDurée_CV ₂	0,15	0,15	0,15	0,15	0,15	0,14
Std Durée_CV ₂	0,02	0,01	0,02	0,02	0,02	0,02

Tableau 6.50 Durée moyenne et écart-type entre la consonne C et la cible voyelle V₂ (Durée_CV₂) dans les neuf séquences pour les six sujets

Observations :

- Le sujet français CV se distingue des autres sujets dans les deux cas : pour les séquences /VtV/ et les séquences /VkV/. Il a des durées les plus courtes que celles des autres sujets.
- Le sujet français AV, dans la plupart des cas dans les séquences /VtV/, se distingue des autres sujets. Elle a des durées plus longues que le sujet CV mais plus courtes que les autres sujets. Pour les séquences /VkV/, elle forme un groupe avec les sujets SK, PB et WS en général.
- Le sujet chinois SK forme parfois un groupe avec les sujets WS et AV dans les séquences /VtV/. Il se distingue des autres dans les cas /atu/, /itu/, /uta/.
- Le sujet chinois WS, le sujet français PB et le sujet chinois JW forment un groupe en général dans tous les cas. Les trois sujets qui ont des durées plus longues.

6.3.2 Analyse de la durée entre la cible voyelle V_1 et la cible voyelle V_2

6.3.2.1 Séquences $/V_1tV_2/$

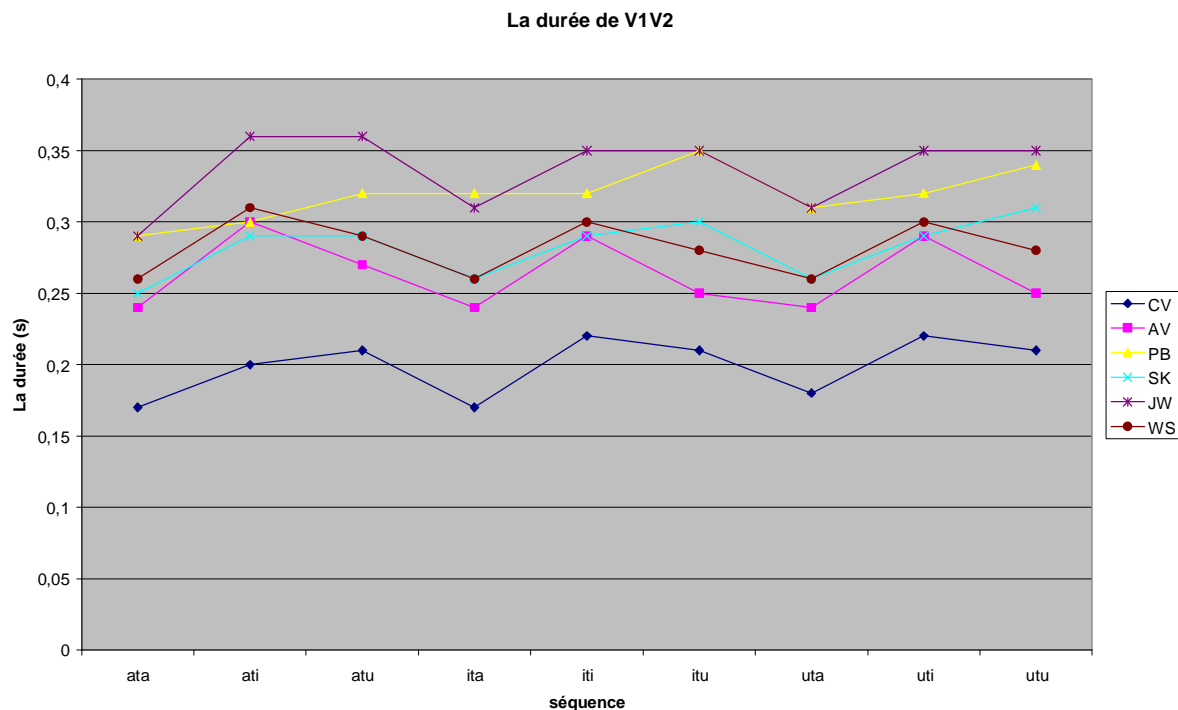


Figure 6.29 Durée moyenne entre la cible voyelle V_1 et la cible voyelle V_2 dans les séquences $/V_1tV_2/$ pour les six sujets.

	ata(s)	ati(s)	atu(s)	ita(s)	iti(s)	itu(s)	uta(s)	uti(s)	utu(s)
Sujet : PB meanDURÉE_ V_1V_2	0,29	0,30	0,32	0,32	0,32	0,35	0,31	0,32	0,34
Sujet : PB stdDURÉE_ V_1V_2	0,01	0,02	0,02	0,03	0,03	0,04	0,03	0,03	0,03
Sujet : CV meanDURÉE_ V_1V_2	0,17	0,20	0,21	0,17	0,22	0,21	0,18	0,22	0,21
Sujet : CV stdDURÉE_ V_1V_2	0,01	0,01	0,01	0,01	0,03	0,02	0,01	0,02	0,02
Sujet : AV meanDURÉE_ V_1V_2	0,24	0,30	0,27	0,24	0,29	0,25	0,24	0,29	0,25
Sujet : AV stdDURÉE_ V_1V_2	0,02	0,04	0,03	0,03	0,04	0,04	0,02	0,04	0,03
Sujet : SK meanDURÉE_ V_1V_2	0,25	0,29	0,29	0,26	0,29	0,30	0,26	0,29	0,31
Sujet : SK stdDURÉE_ V_1V_2	0,03	0,02	0,04	0,03	0,03	0,04	0,03	0,03	0,03
Sujet : JW meanDURÉE_ V_1V_2	0,29	0,36	0,36	0,31	0,35	0,35	0,31	0,35	0,35
Sujet : JW stdDURÉE_ V_1V_2	0,03	0,02	0,07	0,04	0,02	0,05	0,03	0,02	0,04
Sujet : WS meanDURÉE_ V_1V_2	0,26	0,31	0,29	0,26	0,30	0,28	0,26	0,30	0,28
Sujet : WS stdDURÉE_ V_1V_2	0,03	0,04	0,03	0,03	0,03	0,03	0,03	0,03	0,03

Tableau 6.51 Durée moyenne et écart-type entre la cible V_1 et la cible voyelle V_2 (DURÉE_ V_1V_2) dans les séquences $/V_1tV_2/$ pour les six sujets

6.3.2.2 Séquences /V₁kV₂/ :

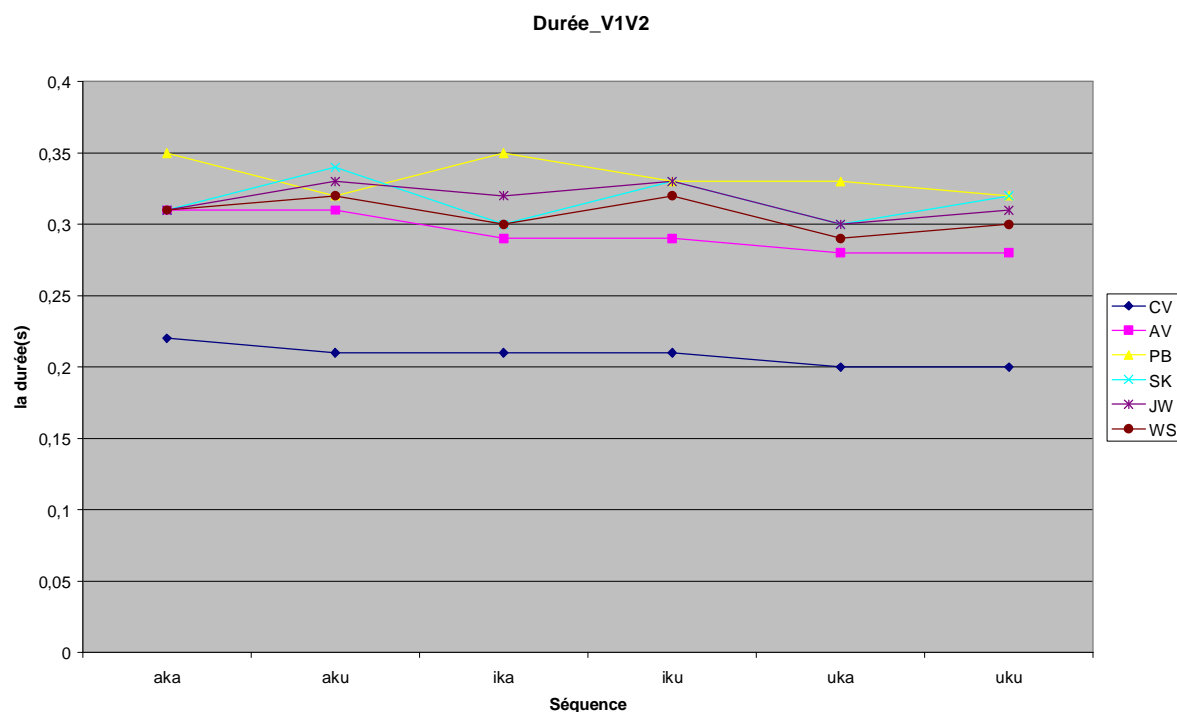


Figure 6.30 Durée moyenne entre la cible voyelle V₁ et la cible voyelle V₂ dans les séquences /V₁kV₂/ pour les six sujets.

	aka(s)	Aku(s)	ika(s)	iku(s)	uka(s)	uku(s)
Sujet : PB						
meanDURÉE_V ₁ V ₂	0,35	0,32	0,35	0,33	0,33	0,32
Sujet : PB						
stdDURÉE_V ₁ V ₂	0,02	0,02	0,02	0,03	0,04	0,03
Sujet : CV						
meanDURÉE_V ₁ V ₂	0,22	0,21	0,21	0,21	0,20	0,20
Sujet : CV						
stdDURÉE_V ₁ V ₂	0,01	0,02	0,02	0,02	0,03	0,03
Sujet : AV						
meanDURÉE_V ₁ V ₂	0,31	0,31	0,29	0,29	0,28	0,28
Sujet : AV						
stdDURÉE_V ₁ V ₂	0,04	0,04	0,04	0,04	0,04	0,04
Sujet : SK						
meanDURÉE_V ₁ V ₂	0,31	0,34	0,30	0,33	0,30	0,32
Sujet : SK						
stdDURÉE_V ₁ V ₂	0,02	0,02	0,02	0,02	0,03	0,03
Sujet : JW						
meanDURÉE_V ₁ V ₂	0,31	0,33	0,32	0,33	0,30	0,31
Sujet : JW						
stdDURÉE_V ₁ V ₂	0,05	0,02	0,03	0,02	0,03	0,03
Sujet : WS						
meanDURÉE_V ₁ V ₂	0,31	0,32	0,30	0,32	0,29	0,30
Sujet : WS						
stdDURÉE_V ₁ V ₂	0,03	0,02	0,03	0,03	0,04	0,04

Tableau 6.52 Durée moyenne et écart-type entre la cible V₁ et la cible voyelle V₂ (DURÉE_V₁V₂) dans les séquences /V₁kV₂/ pour les six sujets

Observations :

- Le sujet français CV se distingue des autres sujets dans les deux séquences /VtV/ et /VkV/. Ses durées sont les plus courtes de tous les sujets.
- Le sujet français AV, le sujet chinois SK et le sujet chinois WS forment un groupe en général dans les séquences /VtV/. Ils ont des durées plus longues que le sujet CV, mais plus courtes que le sujet PB et le sujet JW.
- Le sujet français PB et le sujet chinois JW forment un groupe pour la plupart des cas dans les séquences /VtV/. Ils ont les durées les plus longues des six sujets. Le sujet PB forme un groupe avec WS, SK et AV dans les deux cas /ati/ et /uti/.
- Dans les séquences /VkV/, excepté le sujet CV, tous les sujets forment en général un groupe. Le sujet PB a des durées plus longues pour certain cas, et il se distingue des autres parfois.

Résumé :

- Ü Pour la durée de transition C-V₂ ainsi que pour la durée de transition V₁-V₂, le sujet français CV qui a une durée très courte se distingue des autres dans tous les cas. Le sujet français AV et le sujet chinois SK et WS ne se distinguent pas dans la plupart des cas. Ils ont des durées significativement plus longues que le sujet CV, et plus courtes que les deux autres sujets. Le sujet français PB et le sujet chinois JW ne se distinguent pas significativement dans la plupart des cas. Ils ont des durées plus longues que tous les autres sujets.
- Ü Bien que une durée plus courte est observée chez le sujet français CV et une durée plus longue est observée chez le sujet chinois JW, on ne peut faire de distinction significative entre les deux groupes de langues sur la durée de transition C-V₂ et V₁-V₂.
- Ü Si nous comparons la durée de transition à l'anticipation que nous avons observée pour chaque locuteur, nous trouvons, parmi les sujets français, que le sujet CV qui a une durée la plus courte ne montre pas plus d'anticipation que le sujet PB qui a une durée la plus longue. Nous pouvons dire que, dans nos données, il n'y a pas de relation entre la quantité d'anticipation de V₂ sur V₁, et la durée de transition entre les cibles V₁ et V₂. Nous pouvons rejeter la possibilité d'influence du temps sur la différence d'anticipation observée chez les locuteurs français et chez les locuteurs chinois.

Chapitre 7 : Simulations de séquences VCV

Introduction

Nous rappelons que l'objectif de notre travail est d'étudier l'influence potentielle de la nature des unités phonologiques d'une langue sur la planification de la parole. Nous avons présenté dans le chapitre 2, les différences sur le rôle de la syllabe pour le français et pour le chinois. Dans le chapitre 6, l'analyse de l'anticipation dans ces deux langues montre que, pour les locuteurs français, la voyelle V_2 influence la séquence entière de V_1CV_2 , tandis que cette influence est limitée à la syllabe CV_2 pour les locuteurs chinois. Pour comprendre comment les différentes stratégies pour la planification peuvent intégrer ces différentes contraintes de la structure phonologique, nous avons proposé une modélisation du contrôle moteur dans le chapitre 4. Le présent chapitre a pour but de tester ces hypothèses de planification et d'exécution du mouvement de la parole en utilisant la modélisation que nous avons élaborée. Les questions essentielles sont celles de la nature et de la longueur de séquences prises en compte dans la planification. Pour cela nous avons proposé trois hypothèses de planification : l'hypothèse de voisinage, l'hypothèse syllabique et l'hypothèse inspirée du modèle d'Öhman. Nous nous intéressons également à deux hypothèses d'exécution du mouvement (l'exécution séquentielle et l'exécution selon Öhman). Pour chacune de ces deux exécutions du mouvement, deux hypothèses de planification seront envisagées.

a). Pour l'exécution séquentielle :

- 1) l'hypothèse de planification de voisinage: la planification s'opère sur l'ensemble de la séquence V_1CV_2 et vise à rapprocher au maximum les 3 cibles les une des autres. (Les cibles correspondent aux cibles canoniques de chaque phonème) ;
- 2) l'hypothèse de planification syllabique : la planification s'opère au sein de la syllabe CV_2 ;

b). Pour l'exécution selon Öhman

- 1) l'hypothèse de planification inspirée du modèle d'Öhman : la planification donne à la transition V_1V_2 un statut spécifique;
- 2) l'hypothèse de planification de voisinage : déjà évoquée ci-dessus.

Nous testerons aussi les simulations sans planification (C'est-à-dire fondées sur des séquences de cibles invariantes, quelque soit le contexte) pour faire une comparaison avec celles qui ont été obtenues avec planification. Enfin, nous présentons une évaluation systématique des prédictions dans le plan sagittal pour les séquences VCV. Les voyelles utilisées étaient les

voyelles non arrondies / i, e, ε, a, ɔ, œ / et la consonne était /k/. Les résultats seront comparés avec différentes stratégies observées pour les locuteurs français et les locuteurs chinois dans les données articulatoires.

7.1 Hypothèse de la planification.

Comme nous l'avons dit dans le chapitre 4, la planification de la séquence consiste à déterminer les commandes motrices associées à la génération de caractéristiques spectrales cibles souhaitées pour le signal de la parole. C'est un problème inverse (comment spécifier les commandes du système à partir de la connaissance des objectifs à atteindre). C'est aussi un problème mal posé puisque la solution n'est pas unique (cf. chapitre 3). Pour régler ce problème, nous proposons de minimiser une fonction du coût qui a été élaborée afin de prendre en compte deux critères (voir chapitre 4). Ces deux critères intègrent à la fois des contraintes orientées vers le locuteur (minimisation d'un « effort ») et des contraintes orientées vers l'auditeur (perception des sons successifs).

Les hypothèses de planification se différencient l'une de l'autre dans le calcul de la fonction de coût. Nous envisagerons ainsi différentes hypothèses sur la longueur et la nature de la séquence sur laquelle le coût est calculé. La fonction de coût correspond à une certaine prise en compte de l'effort du locuteur. Nous avons choisi de caractériser cet effort par le chemin parcouru pendant la séquence par les commandes cibles dans l'espace des commandes musculaires. La fonction de coût est ainsi définie comme la somme pondérée des distances entre phonèmes successifs d'une séquence dans cet espace. La pondération des distances varie avec l'importance que chaque transition entre phonèmes prend dans la phonologie de la langue étudiée. Cette fonction de coût prend en compte aussi les contraintes perceptives. Autrement dit, ce coût perceptif a une valeur faible dans l'espace acoustico-perceptif pour tous les sons à l'intérieur de leur région cible. Dès qu'on passe à l'extérieur de la région cible, la fonction de coût augmente brutalement.

7.1.1 Hypothèse de planification de voisinage

L'hypothèse de planification de voisinage vise à minimiser la taille du voisinage défini par les trois cibles de séquence VCV. Elle prend en compte tous les phonèmes de manière équivalente. La fonction de coût est définie comme $d(V_1, C) + d(C, V_2) + d(V_2, V_1)$. La figure 7.1 schématise la façon de cette hypothèse de type voisinage.

La figure 7.2 présente un exemple de résultat obtenu dans l'espace de sortie (F_1 , F_2) avec

l'hypothèse de voisinage pour la planification de la séquence /ake/. Les ellipses représentent les trois régions cibles des sons élémentaires /a/, /k/, /e/ de la séquence considérée. Les lignes droites qui relient ces ellipses, connectent entre elles les positions cibles associées, sélectionnées par le processus de planification optimale. (Voir le chapitre 4 pour l'algorithme d'optimisation).

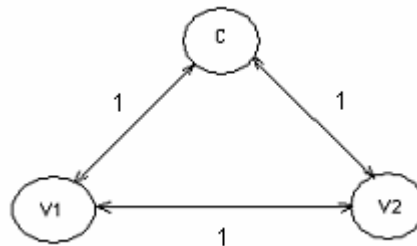


Figure 7.1 Hypothèse de planification de voisinage : toutes les transitions entre phonèmes de la séquence VCV sont équivalentes pour la planification. (Pondération égale à 1).

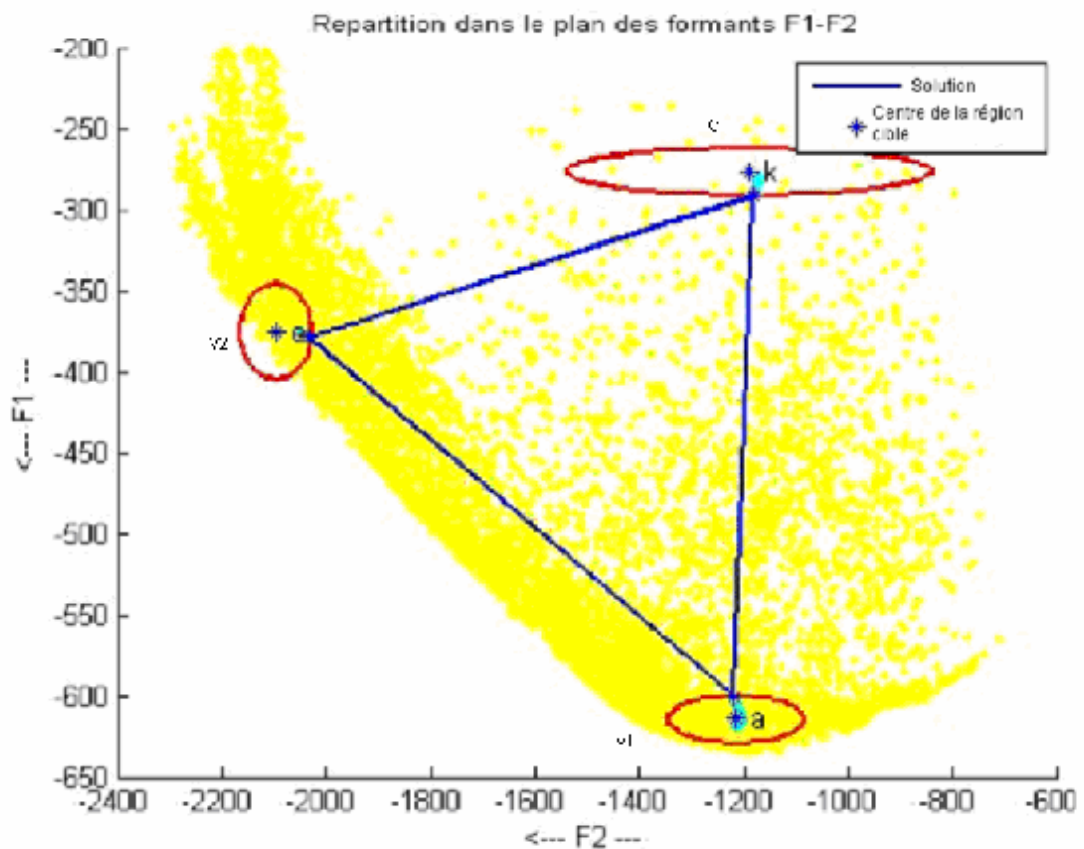


Figure 7.2 Résultats dans le plan (F₁, F₂) de la planification d'une séquence /ake/ selon l'hypothèse de planification de voisinage.

7.1.2 Hypothèse de planification syllabique

L'hypothèse de planification syllabique est une planification qui donne un statut spécifique au voisinage de la syllabe CV₂ dans la planification. La figure 7.3 schématise l'hypothèse syllabique.

La figure 7.4 présente un exemple de résultat dans le plan (F₁, F₂) pour la planification de la séquence /ake/, obtenu avec l'hypothèse de planification syllabique.

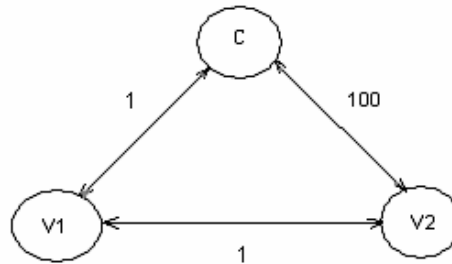


Figure 7.3 Hypothèse de planification syllabique : la transition CV₂ a un poids 100 fois plus important que les autres dans le calcul de coût.

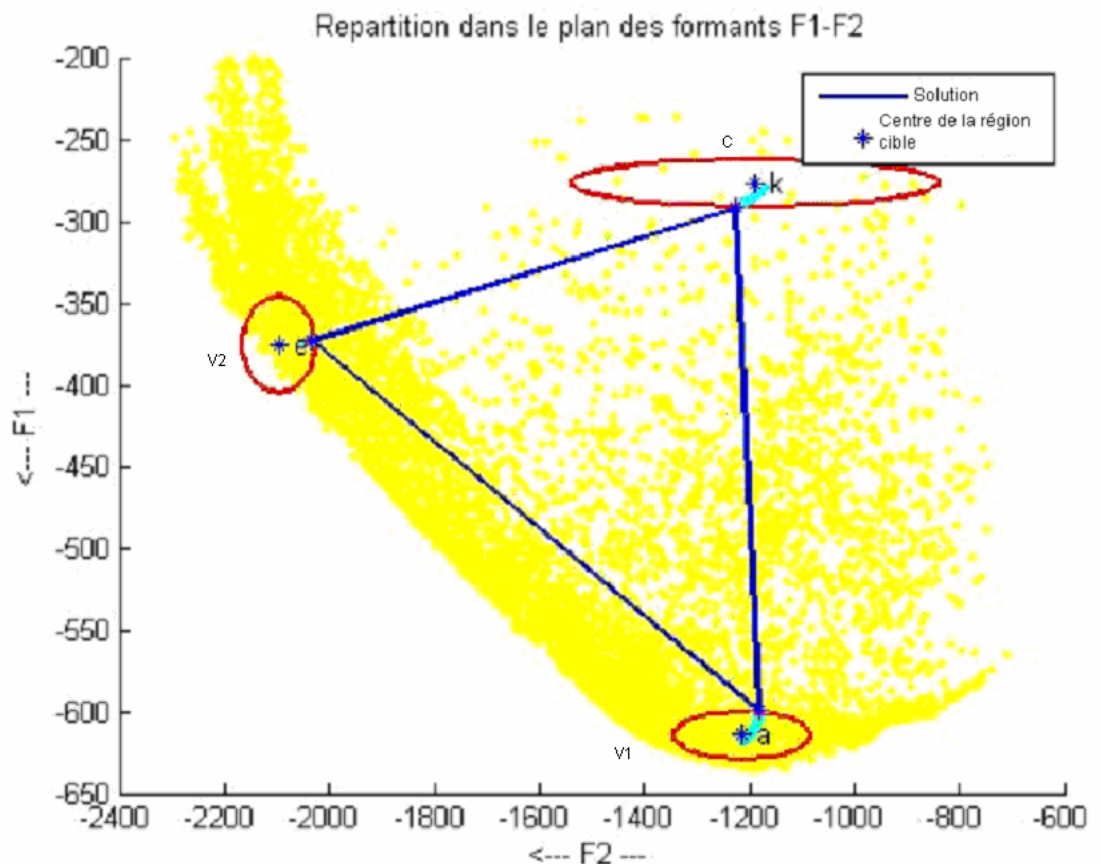


Figure 7.4 Résultats dans le plan (F₁, F₂) de la planification d'une séquence /ake/ selon l'hypothèse de planification syllabique.

Si nous comparons les deux figures 7.2 et 7.4, nous observons que la voyelle V_1 (/a/) pour l'hypothèse de planification de voisinage (figure 7.2) est plus proche (plus à gauche) des deux autres phonèmes que pour l'hypothèse de planification syllabique (figure 7.4). Par contre, nous observons que la consonne /k/ est plus proche de la voyelle V_2 /e/ pour l'hypothèse de planification syllabique que pour l'hypothèse de planification de voisinage.

7.1.3 Hypothèse de planification inspirée du modèle d'Öhman

L'hypothèse de planification inspirée du modèle d'Öhman est une planification qui donne un statut spécifique à la transition vocalique V_1V_2 . L'objectif est de minimiser ces deux cibles. La figure 7.5 présente le schéma de cette hypothèse.

La figure 7.6 présente un exemple de résultat de cette planification pour la séquence /ake/. Si nous comparons les figures 7.2 et 7.6, nous observons un effet clair sur la voyelle V_1 /a/ et sur la voyelle V_2 /e/ pour l'hypothèse selon Öhman que pour l'hypothèse de voisinage. La voyelle /a/ a choisi son chemin dès son départ vers la voyelle V_2 /e/. La voyelle /e/ se rapproche ainsi de la première voyelle /a/. La consonne /k/ est définie indépendamment de la transition vocalique.

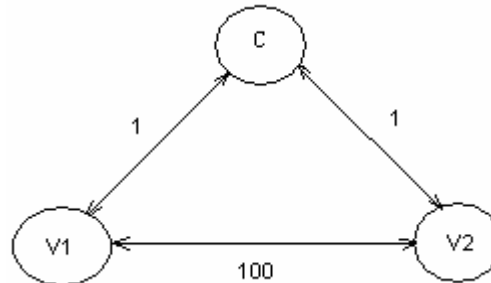


Figure 7.5 Hypothèse de planification inspirée du modèle d'Öhman : la transition V_1V_2 a un poids 100 fois plus important que les autres dans le calcul de coût.

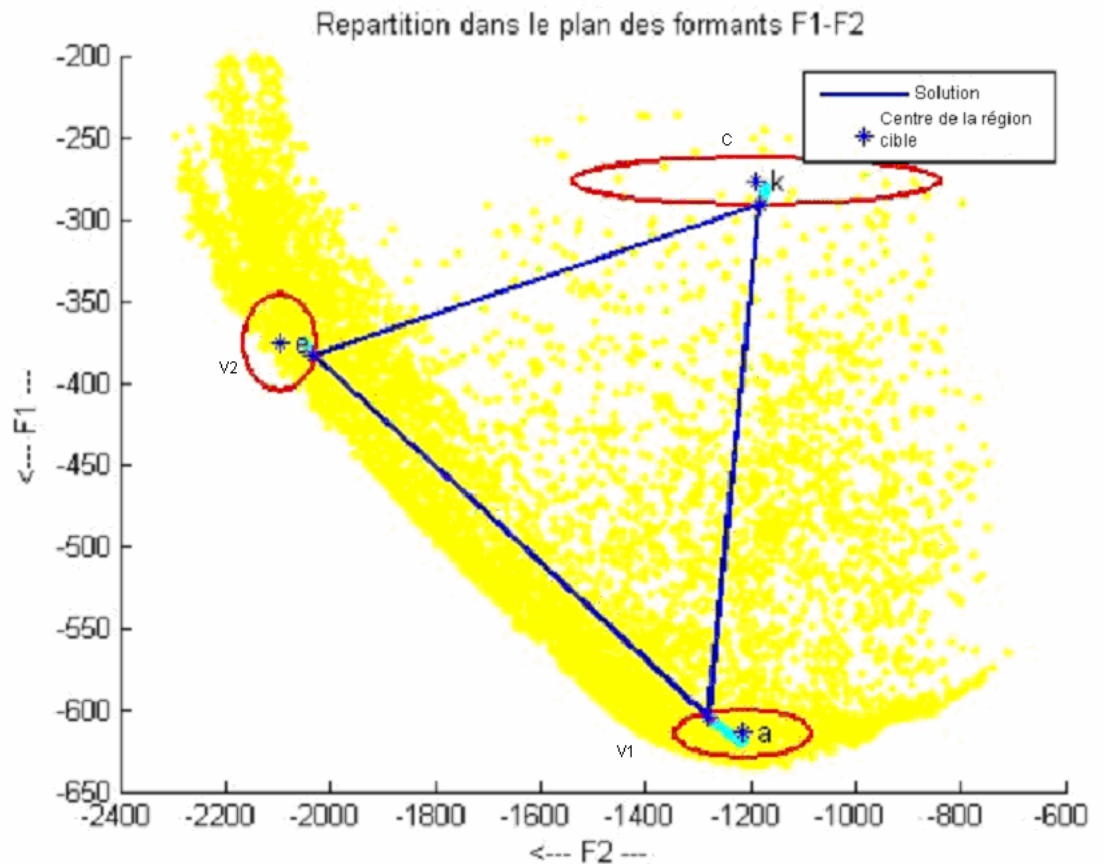


Figure 7.6 Résultats dans le plan (F_1 , F_2) de la planification d'une séquence /ake/ selon l'hypothèse de planification selon Öhman.

Nous présentons dans les figures 7.7 et 7.8, le résultat de trois hypothèses de planification par le contour de la langue planifié à la cible pour $V_1=/a/$ et pour $C=/k/$. Pour mesurer quantitativement l'effet des différentes hypothèses, nous présentons également un contour de la langue "référence" correspondant au point de départ de l'algorithme (centre d'ellipse caractérisant la région cible de chaque phonème dans les plans F_1 - F_2 et F_2 - F_3 . Pour le principe de l'algorithme voir chapitre 4)

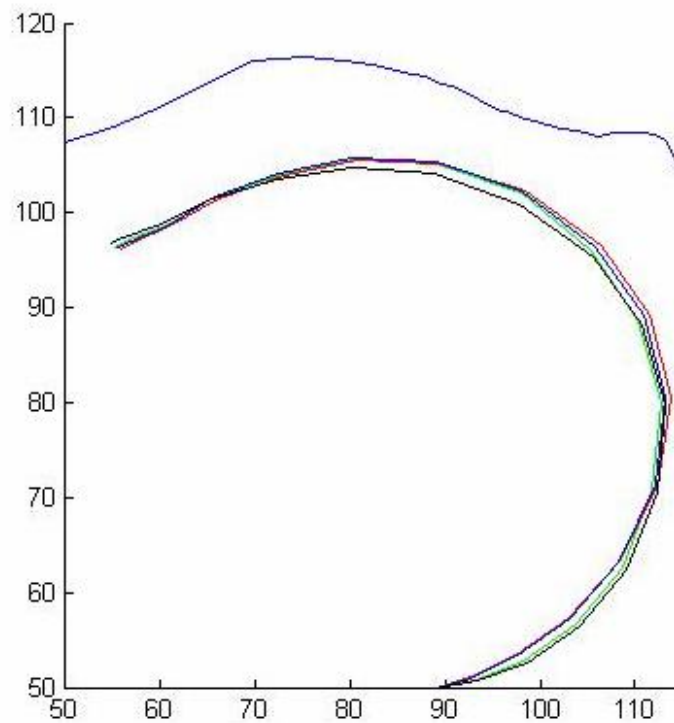


Figure 7.7 Contour de la langue cible planifié de $V_1=/a/$ dans la séquence /ake/ pour trois hypothèses de planification. Trait rouge : hypothèse de planification syllabique, trait bleu : hypothèse de planification de voisinage et trait vert : hypothèse de planification selon Öhman. Le contour de langue “référence” est présenté en noir.

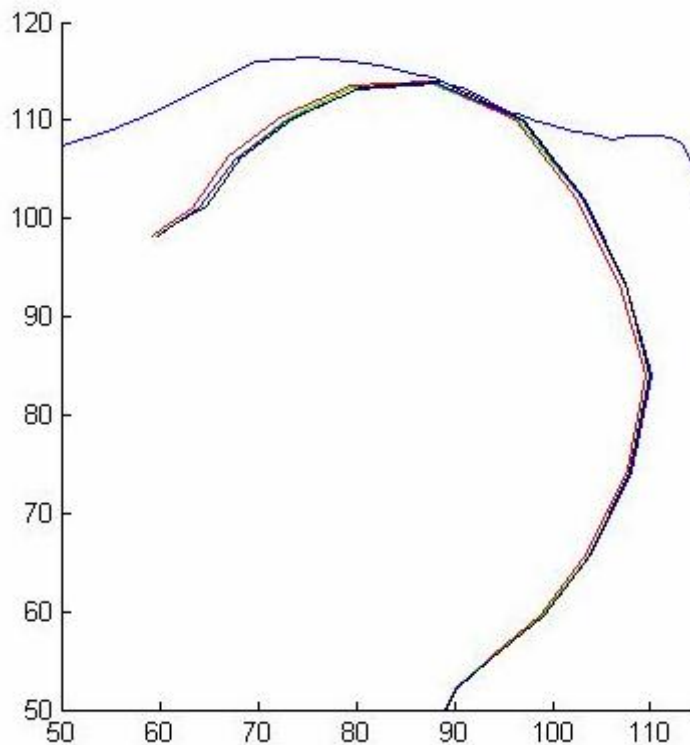


Figure 7.8 Contour de la langue cible planifié de $C=/k/$ dans la séquence /ake/ pour trois hypothèses de planification. Trait rouge : hypothèse de planification syllabique, trait bleu : hypothèse de planification de voisinage et trait vert : hypothèse de planification selon Öhman. Le contour de langue “référence” est présenté en noir.

Nous observons dans la figure 7.7 que le contour de langue cible planifié pour $V_1=/a/$ est plus antérieur pour l'hypothèse inspirée du modèle d'Öhman que pour les deux autres hypothèses. C'est-à-dire que cette hypothèse montre plus d'influence de l'anticipation de voyelle $V_2=/e/$ sur V_1 . Cela est conforme avec le fait que l'hypothèse inspirée du modèle d'Öhman donne un statut spécifique à la transition vocalique. L'hypothèse syllabique montre le contour de langue cible le plus postérieur parmi ces trois hypothèses. Cette observation est cohérente car l'hypothèse syllabique donne un statut spécifique à la syllabe et par conséquent la voyelle postérieure V_1 est peu influencée par la voyelle antérieure V_2 . L'hypothèse de voisinage propose un résultat intermédiaire. La position de $V_1=/a/$ est moyennement influencée par la deuxième voyelle $/e/$. L'amplitude maximale des écarts entre les cibles planifiées par ces trois hypothèses est inférieure à 2mm, mais elle se situe juste dans la partie de la constriction de la voyelle $/a/$. Le contour de langue "référence" est plus bas que les cibles résultant des trois hypothèses de planification. Ceci est conforme avec le contexte droit constitué d'une voyelle haute antérieure $/e/$ et une consonne palatale $/k/$.

La figure 7.8 montre que la langue de la consonne $/k/$ est plus antérieure pour l'hypothèse syllabique que pour les deux autres hypothèses. La consonne est ainsi plus affectée par la deuxième voyelle $V_2 /e/$ pour l'hypothèse syllabique. Cela est conforme avec la principale de base de cette hypothèse. Les autres hypothèses ne montrent pas beaucoup de différence pour la position de la consonne.

7.2 Simulation des séquences VCV

Pour tester ces hypothèses de planification et en évaluer quantitativement les différences, nous avons effectué des simulations de type VCV. Le temps de transition entre les positions cibles d'une séquence est égal à 40ms et le temps de tenue des cibles à 100ms. Nous avons visualisé quelques résultats en observant la position du nœud du contour supérieur de la langue qui est la plus proche de la position du capteur T_4 utilisé dans les expériences EMMA (cf. figure 7.9).

Rappelons qu'une fois les cibles planifiées, l'exécution de mouvement pourra se faire selon deux hypothèses. La première de ces hypothèses est l'exécution séquentielle qui prend en compte tous les phonèmes de manière équivalente dans l'exécution du mouvement articulaire. La deuxième est l'exécution selon Öhman qui donne à la transition vocalique un statut tout particulier dans l'exécution du mouvement. Pour chacune de ces deux exécutions, nous testons deux hypothèses de planification.

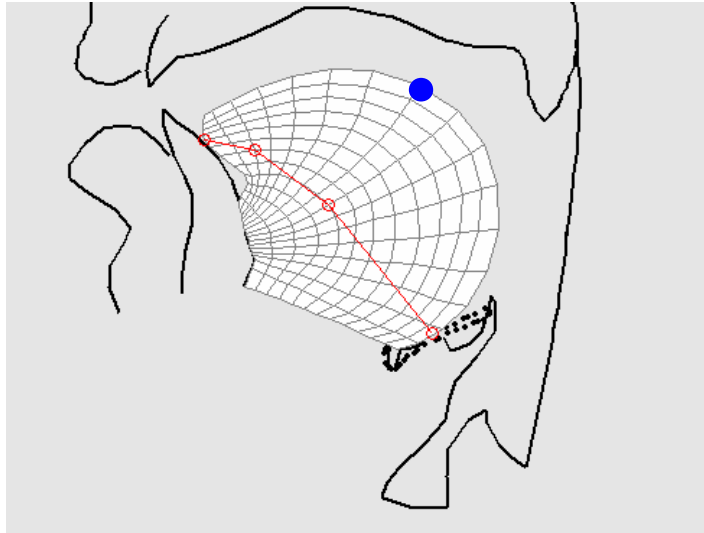


Figure 7.9 : Présentation du point sur le contour de la langue (cercle plein) dont les mouvements sont représentés dans les figures suivantes.

7.2.1 L'exécution séquentielle du mouvement

La figure 7.10 présente le schéma temporel de principe de la variation des commandes motrices pour l'exécution séquentielle du mouvement pour la séquence /a-k-e/. Les temps des transitions vers les cibles sont de 40ms. Les temps de tenue pour toutes les cibles sont de 100ms. La langue est au départ dans sa position de repos. L'axe des temps est donnée en milliseconde. L'axe vertical représente seulement un état à atteindre pour chaque phonème. Par exemple, un état $y=1$ pour atteindre la voyelle /a/ et un état $y=3$ pour atteindre la /k/.

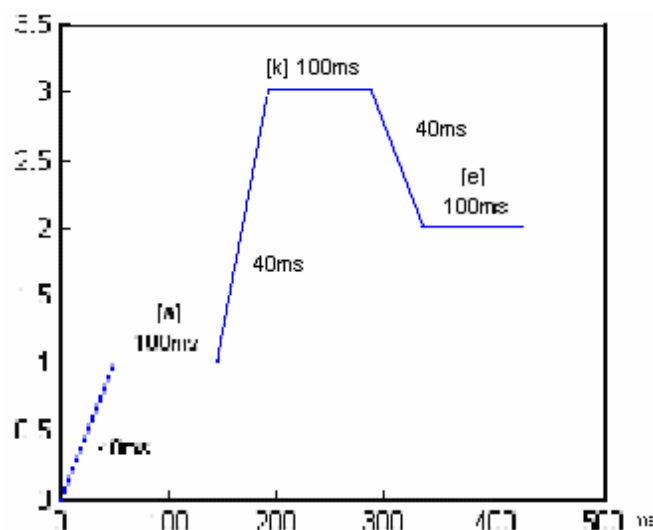


Figure 7.10 Schéma temporel (mm) pour la variation des commandes motrices pour l'exécution séquentielle pour la séquence /repos-a-k-e/.

La figure 7.11 présente les trajectoires correspondantes des commandes motrices. Les commandes motrices des muscles (GGP, GGA, HYO, STY, VER, SL, IL) sont présentées en mm.

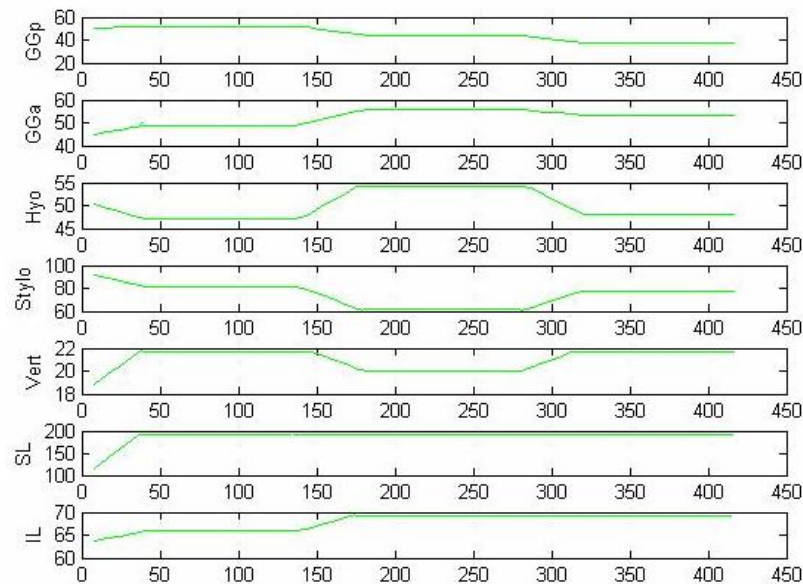


Figure 7.11 Evolution des commandes motrices pour l'exécution séquentielle pour la séquence /repos-a-k-e/ (Commandes en mm et temps en ms)

7.2.1.1 Les simulations pour une planification de type voisinage avec exécution séquentielle

La figure 7.12 montre l'évolution temporelle de la position et de la vitesse tangentielle du nœud considéré sur le contour de la langue (cf. figure 7.9) pour l'hypothèse de voisinage et pour les séquences /a-k-ɔ/ et /a-k-e/.

La figure (a) présente le déplacement de la langue suivant la direction X, et la figure (b) présente le déplacement de la langue suivant la direction Y. La figure (c) présente la vitesse tangentielle.

La langue est au départ dans sa position de repos. La langue recule (x croît) et descend un peu (y décroît) vers la voyelle /a/ qui à son tour est maintenue pendant le temps de tenue. La langue monte ensuite (y croît) pour aller produire la consonne /k/ qui est marquée par un long plateau selon la direction Y. Ce plateau est dû au contact entre la langue et le palais qui limite l'ascension de la langue vers sa cible virtuelle. Puis, la langue va avancer et s'élever vers /e/ ou reculer et s'abaisser vers /ɔ/.

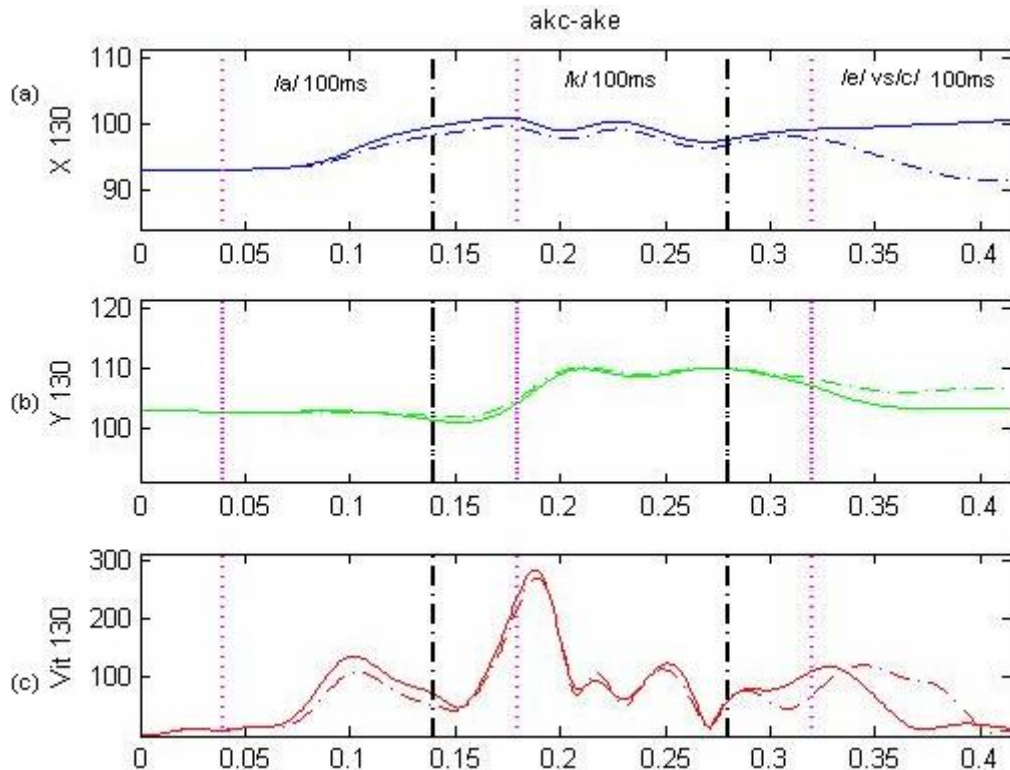


Figure 7.12 Déplacements horizontaux (a), et verticaux (b), et vitesses tangentielles (c) du point du contour supérieur de la langue donnée en figure 7.9 pour l'hypothèse de planification de voisinage avec exécution séquentielle. /a-k-ɔ/ trait continu ; /a-k-e/ trait tiret. Les traits violets verticaux représentent le début de la commande de tenue de chaque phonème. Les traits noirs verticaux représentent la fin de la commande de tenue de chaque phonème.

Nous observons que les positions cibles des différents sons sont donc bien atteintes. La séparation des signaux de position existe dès la tenue de voyelle V_1 . La position de V_1 =/a/ et la position de /k/ sont affectées par la nature de la voyelle V_2 . La vitesse tangentielle est affectée juste après la transition /repos/-/a/.

7.2.1.2 Les simulations pour une planification de type syllabique avec exécution séquentielle

La figure 7.13 montre les simulations pour l'hypothèse syllabique. Elle présente les déplacements horizontaux et verticaux, ainsi que les vitesses tangentielles du point de contour supérieur du modèle représenté à la figure 7.9 pour les séquences /a-k-ɔ/ (trait continu) et /a-k-i/ (trait tiret).

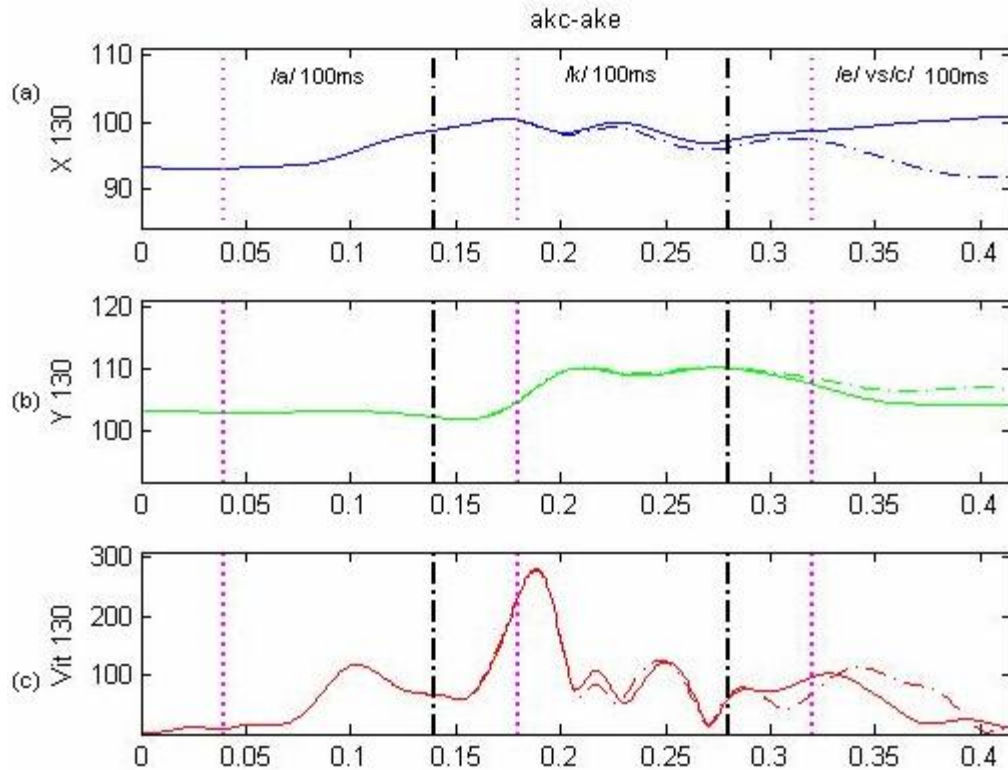


Figure 7.13 Déplacements horizontaux (a), et verticaux (b), et vitesses tangentielles (c) du point du contour supérieur de la langue donnée en figure 7.9 pour l'hypothèse de planification syllabique avec exécution séquentielle. /a-k-ɔ/ trait continu ; /a-k-e/ trait tiret. Les traits violets verticaux représentent le début de la commande de tenue de chaque phonème. Les traits noirs verticaux représentent la fin de la commande de tenue de chaque phonème.

Nous observons que la séparation des signaux de position existe dans la tenue consonantique avant que la commande ne spécifie la transition vers la voyelle V_2 , juste après le début du contact consonantique entre la langue et le palais. D'où l'apparition d'un effet d'anticipation après la transition de la voyelle /a/ vers la consonne /k/. La position de la voyelle /a/ n'est pas affectée par la voyelle V_2 ; par contre on observe un effet clair d'anticipation pour la consonne /k/. Le profil de vitesse est également affecté par ce phénomène d'anticipation, mais pas dans la transition de la voyelle /a/ vers la consonne.

7.2.1.3 Les simulations sans planification avec exécution séquentielle

La figure 7.14 montre les simulations sans planification. Elle présente l'évolution temporelle de la position et de la vitesse tangentielle du nœud considéré sur le contour de la langue (cf. figure 7.9) pour les séquences /a-k-ɔ/ et /a-k-e/.

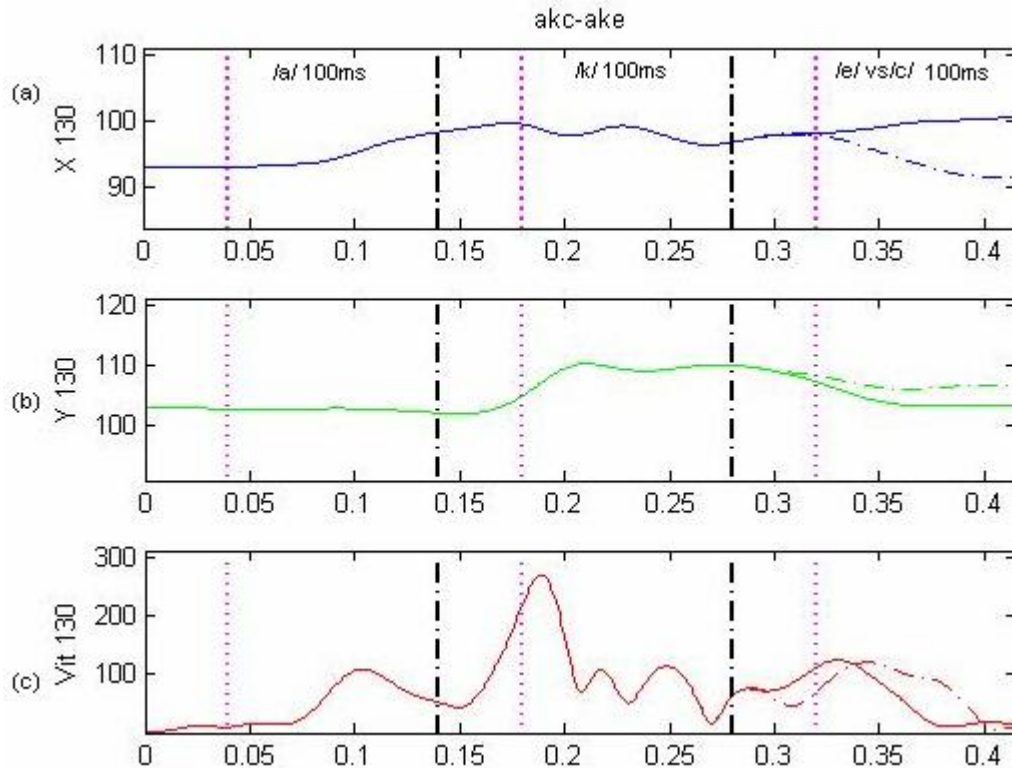


Figure 7.14 Déplacements horizontaux (a), et verticaux (b), et vitesses tangentielles (c) du point du contour supérieur de la langue donnée en figure 7.9 pour l'hypothèse sans planification avec exécution séquentielle. /a-k-ɔ/ trait continu ; /a-k-e/ trait tiret. Les traits violets verticaux représentent le début de la commande de tenue de chaque phonème. Les traits noirs verticaux représentent la fin de la commande de tenue de chaque phonème.

L'hypothèse sans planification donne des résultats pas très réalistes, car elle n'intègre aucune prise en compte du contexte. Elle permet cependant de mesurer d'éventuels effets de d'anticipation physique qui pourraient être liés aux caractéristiques inertielles des articulateurs. La position atteinte pour la consonne [k] et la position atteinte pour la première voyelle /a/ ne sont pas affectées par la voyelle V_2 . Les signaux du mouvement se séparent à la fin de la transition de /k/- V_2 . Le profil de vitesse est affecté un peu plus que les mouvements en X et en Y, mais aussi strictement limité à la transition /k/- V_2 . L'hypothèse sans planification nous montre donc que dans les conditions de timing proposées le modèle purement physique ne génère pas de phénomène d'anticipation.

7.2.2 L'exécution du mouvement selon Öhman

La figure 7.15 présente le schéma temporel de principe de la variation des commandes motrices pour l'exécution du mouvement selon Öhman pour la séquence /repos-a-k-e/. L'hypothèse selon Öhman donne à la transition vocalique un statut tout particulier dans l'exécution du mouvement. La consonne sera considérée comme une perturbation superposée à la transition vocalique. Les temps de tenue pour les deux voyelles sont de 100ms. Le temps de cette transition vocalique est de 180ms (trait noir dans la figure 7.15). Il est exactement égal à l'intervalle entre la fin de la voyelle V_1 et le début de la voyelle V_2 dans l'exécution séquentielle. Pour définir le temps de tenue et le temps de transition de la consonne, nous chercherons à respecter deux contraintes : 1) de maintenir un temps de tenue consonantique équivalente à celui de l'exécution séquentielle. 2) de réduire le temps de transition pour intégrer l'hypothèse de "perturbation" associé à la consonne, tout en restant dans un intervalle raisonnable pour des gestes de parole. Donc, le temps de tenue consonantique reste à 100ms. Les temps des transitions superposées à la base vocalique vers la cible consonne sont de 25ms. L'instant de départ de la consonne sur la base vocalique est situé 15ms après le début de la transition vocalique.

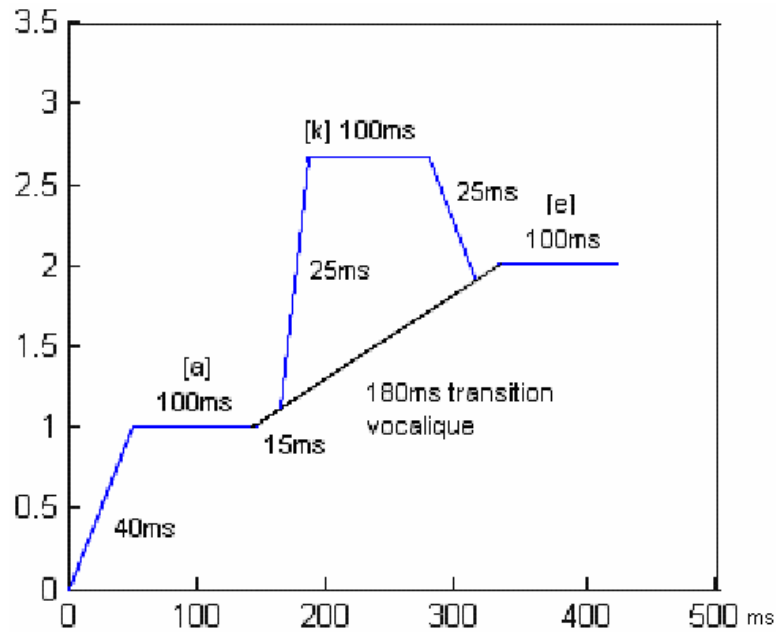


Figure 7.15 Schéma temporel du type de variation des commandes motrices pour l'exécution selon Öhman pour la séquence /repos-a-k-e/. Le trait noir représente le temps de transition vocalique qui est de 180ms. L'instant de départ de la consonne sur cette base vocalique est situé 15ms après le début de la transition vocalique.

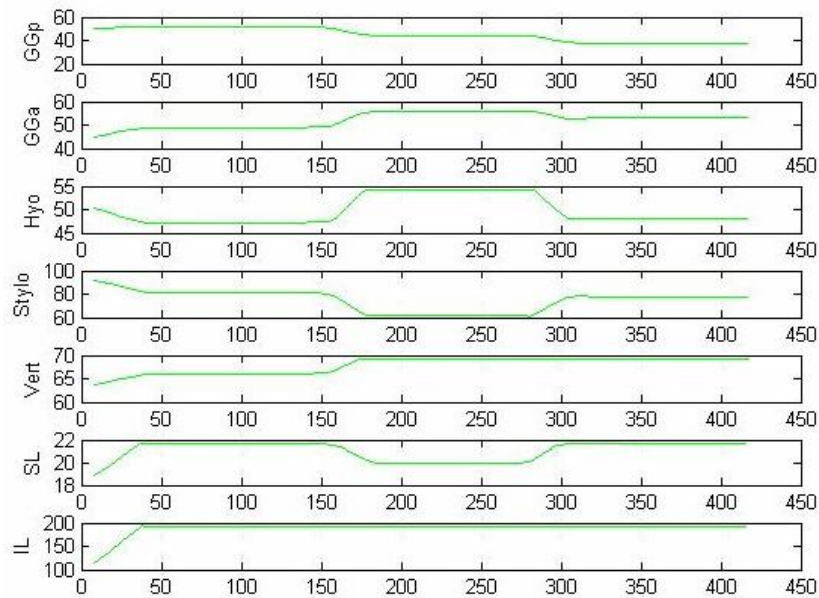


Figure 7.16 Evolution des commandes motrices pour la séquence /repos-a-k-e/ pour l'exécution selon Öhman. (Commandes en mm et temps en ms)

7.2.2.1 Les simulations pour une planification selon Öhman avec exécution selon Öhman

Le figure 7.17 montre l'évolution temporelle de la position et de la vitesse tangentielle du nœud considéré sur le contour de la langue (cf. figure 7.9) pour l'hypothèse de planification selon Öhman pour les séquences /a-k-ɔ/ et /a-k-e/.

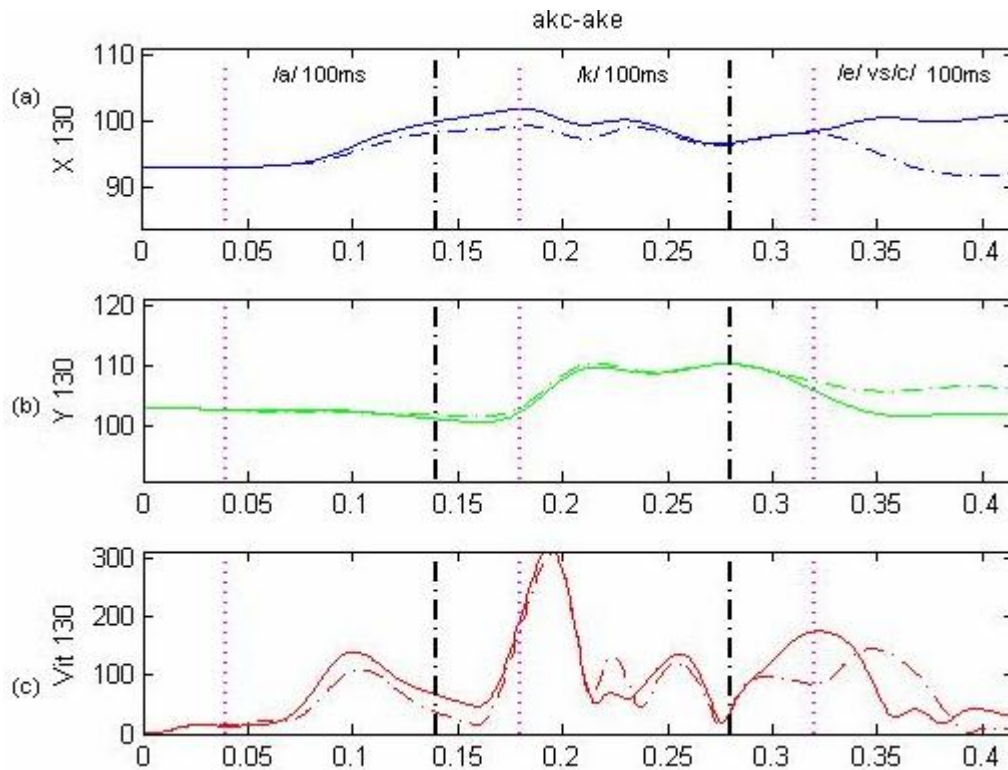


Figure 7.17 Déplacements horizontaux (a), et verticaux (b), et vitesses tangentielles (c) du point du contour supérieur de la langue donnée en figure 7.9 pour l'hypothèse de planification selon Öhman avec exécution selon Öhman. /a-k-ɔ/ trait continu ; /a-k-e/ trait tiret. Les traits violets verticaux représentent le début de la commande de tenue de chaque phonème. Les traits noirs verticaux représentent la fin de la commande de tenue de chaque phonème.

L'hypothèse de planification selon Öhman associé au schéma d'exécution selon Öhman montre un large effet d'anticipation de la voyelle V_2 sur la position de la voyelle V_1 /a/. Par contre cet effet est plus petit sur la position atteinte de /k/. Cependant, la première partie de la tenue consonantique est très différente selon la direction X. Ce n'est qu'à la fin de la tenue que les deux trajectoires se rejoignent. Donc la voyelle V_1 /a/ est beaucoup plus influencée que la consonne /k/. La vitesse tangentielle est affectée juste après la transition de /repos/ vers /a/. Le début du mouvement depuis la voyelle /a/ commence plus tôt pour la séquence /akɔ/ que pour la séquence /aki/.

7.2.2.2 Les simulations pour une planification de type voisinage avec exécution selon Öhman

La figure 7.18 montre les évolutions temporelles de la position et des vitesses tangentielles du nœud considéré sur le contour de la langue (cf. figure 7.9) pour l'hypothèse de planification de voisinage pour les séquences /a-k-ɔ/ et /a-k-e/.

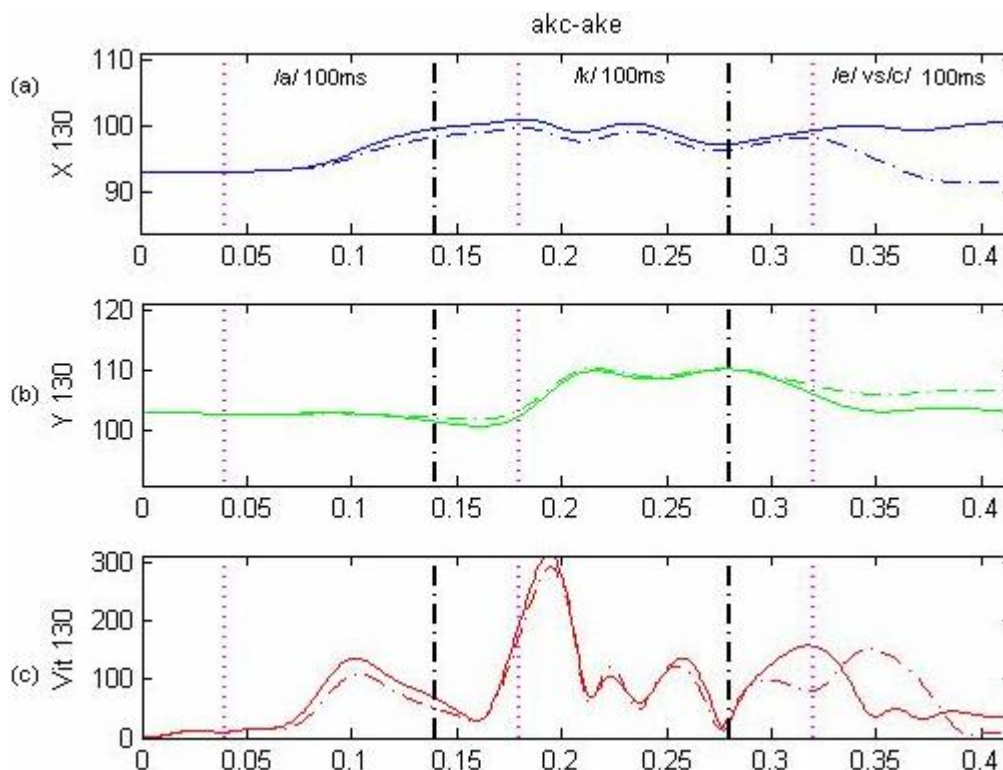


Figure 7.18 Déplacements horizontaux (a), et verticaux (b), et vitesses tangentielles (c) du point du contour supérieur de la langue donnée en figure 7.9 pour l'hypothèse planification de voisinage avec exécution selon Öhman. /a-k-ɔ/ trait continu ; /a-k-e/ trait tiret. Les traits violets verticaux représentent le début de la commande de tenue de chaque phonème. Les traits noirs verticaux représentent la fin de la commande de tenue de chaque phonème.

Nous avons observé que le résultat pour l'hypothèse de planification de voisinage avec exécution selon Öhman n'est pas très différent de celui qui a été obtenu pour l'hypothèse de voisinage avec exécution séquentielle (voir figure 7.12) sur les positions atteintes pour la voyelle V_1 et pour la consonne /k/. Par contre, nous avons observé un décalage temporel du mouvement. Le premier passage par zéro de la vitesse pour la transition de /a/ vers /k/ a environ 15 ms de retard (à l'instant de 0.165) par rapport l'exécution séquentielle (à 0.15). Il est cohérent avec le fait que dans l'hypothèse d'exécution selon Öhman la transition vers la consonne commence 15 ms après la tenue de la voyelle /a/.

7.2.2.3 Effet de décalage de l'initiation du geste consonantique

Les résultats présentés ci-dessus suggèrent que pour une même planification, les deux modèles d'exécution ne gèrent pas de différence sur les positions atteintes par la langue. Deux explications peuvent être proposées pour cette absence de différence : 1) les temps que nous avons choisis pour les temps de tenue et pour les temps de transition sont suffisamment longs pour que la langue atteigne les positions cibles planifiées, ou dans les cas possibles, la position la plus proche possible de la cible planifiée. 2) le décalage entre la fin de la commande de voyelle V_1 et le début du geste consonantique est insuffisant pour qu'un mouvement de V_1 à V_2 soit effectivement perceptible. Pour ne pas cumuler les effets de l'exécution gestuelle avec des problèmes ratage potentiels de cible, nous n'avons pas modifié les durées de tenue et de transition. Par contre, nous avons testé l'effet d'un décalage important de l'initiation du geste consonantique dans l'exécution selon Öhman en fixant ce décalage à 50ms.

La figure 7.19 présente le schéma temporel de principe de la variation des commandes motrices pour cette exécution à la Öhman pour la séquence /repos-a-k-e/. Le temps de la transition vocalique est de 250ms (trait noir dans la figure 7.19). Les temps des transitions de consonne restent de 25ms. Le temps de départ de cette transition de la cible consonne est à 50 ms après la tenue de voyelle /a/. En comparaison avec l'exécution séquentielle, les temps de tenue pour toutes les cibles restent à 100ms. La langue est au départ dans sa position de repos.

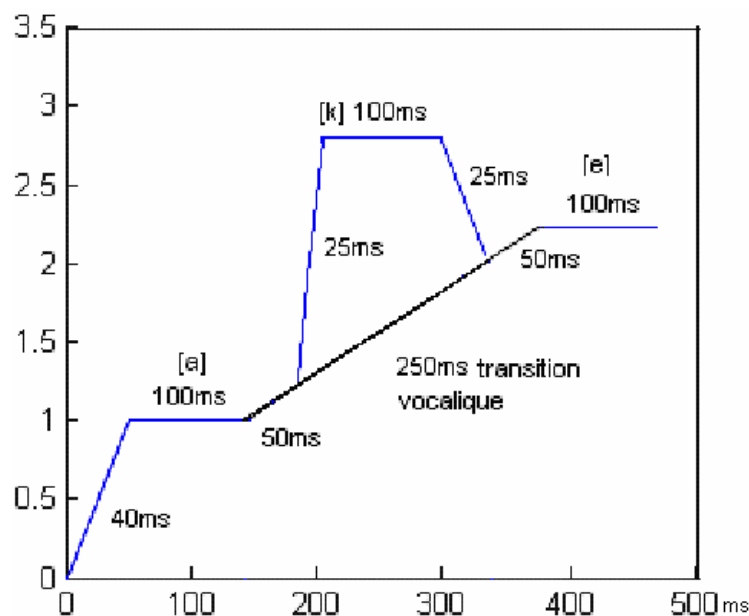


Figure 7.19 Schéma temporel du type de variation des commandes motrices pour l'exécution selon Öhman pour la séquence /repos-a-k-e/. Le trait noir représente le temps de transition vocalique qui est de 250ms. L'instant de départ de la consonne sur cette base vocalique est situé 50ms après le début de la transition vocalique.

Nous observons les conséquences de ce décalage dans la figure 7.20 : un décalage temporel d'environ 50ms est visible sur les signaux de mouvement et de vitesse après la tenue de la voyelle /a/ (à 190 ms). La position atteinte pour /a/ est également influencée par cette exécution, si nous la comparons avec l'exécution séquentielle (la figure 7.12). A 190 ms, nous observons en effet une influence de V_2 sur la position verticale de la langue avant le déclenchement du geste vers la consonne qui montre que la langue commence par se déplacer vers la voyelle V_2 . La différence en position horizontale est moins forte. Dans ces conditions donc, l'exécution selon Öhman tend à augmenter l'effet d'anticipation avant la production de la consonne. Cependant, il convient de noter qu'un tel décalage dans le déclenchement du geste consonantique ne semble pas très réaliste. Ceci suggère donc que le modèle d'exécution selon Öhman n'est pas susceptible de se différencier d'un modèle d'exécution séquentiel dans des conditions de production.

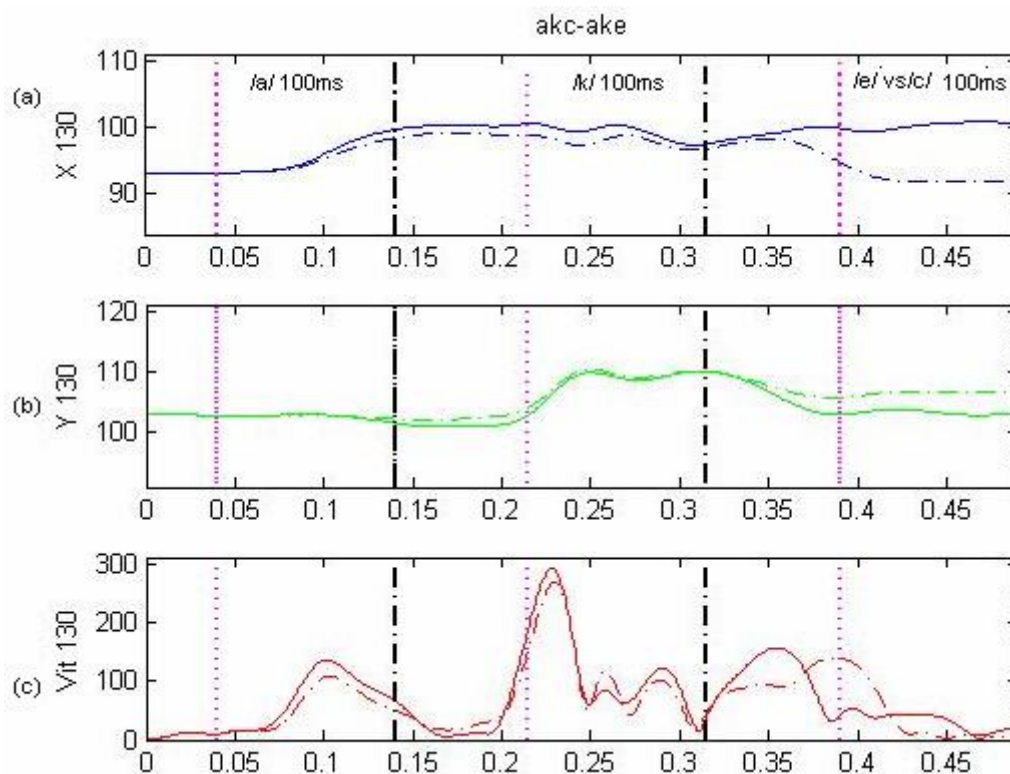


Figure 7.20 Déplacements horizontaux (a), et verticaux (b), et vitesses tangentielles (c) du point du contour supérieur de la langue donnée en figure 7.9 afin de tester un effet de décalage de l'initiation du geste consonantique. /a-k-ɔ/ trait continu ; /a-k-e/ trait tiret. Les traits violets verticaux représentent le début de la commande de tenue de chaque phonème. Les traits noirs verticaux représentent la fin de la commande de tenue de chaque phonème.

7.3 Evaluation systématique des simulations dans le plan sagittal

Nous présentons dans cette section une évaluation systématique des simulations dans le plan sagittal pour des séquences V_1CV_2 . Les voyelles utilisées étaient celles que nous avons présentées dans le chapitre 4: / i, e, ε, a, ɔ, œ /. La consonne était la consonne /k/. Nous avons fait 10 simulations pour chaque séquence V_1CV_2 . Dans notre modélisation du contrôle, la variabilité des réalisations d'une même séquence, d'une répétition à l'autre, est générée de la manière suivante. Au début de l'algorithme d'optimisation, les commandes motrices initiales utilisées dans le processus d'optimisation de l'erreur ont été choisies aléatoirement à l'intérieur de l'ellipse cible, à proximité de son centre.

Pour comparer les résultats de nos simulations avec les données articulatoires que nous avons analysées (chapitre 6), nous présentons ici 4 points du contour supérieur du modèle de la langue dans le plan sagittal, qui correspondent sensiblement aux positions des 4 capteurs (T_1, T_2, T_3, T_4) des données réelles. L'anticipation de la voyelle V_2 sur la voyelle précédente V_1 et l'anticipation de la voyelle V_2 sur la consonne /k/ seront évaluées pour nos différentes hypothèses. Les positions atteintes par la langue pour la voyelle V_1 et pour la consonne /k/ sont calculées à l'instant de la fin de leur tenue. Nous nous concentrons sur trois modélisations de contrôle: (1) planification de voisinage avec exécution séquentielle. (2) planification syllabique avec exécution séquentielle. (3) planification inspirée du modèle d'Öhman avec exécution selon Öhman. Les simulations obtenues avec les deux autres modèles s'appuyant sur des hypothèses de type voisinage avec exécution selon Öhman (voir le principe pour Figure 7.15 et Figure 7.19) seront également présentées en fin de chapitre.

Conformément à ce qui a été fait pour les données expérimentales, nous avons fait des analyses de variance de type ANOVA à Mesures Répétées pour ces trois premières modèles. Les variables dépendantes étaient les positions horizontales et verticales des 4 points du contour supérieur de la langue de V_1 et de /k/, et la variable indépendante était la nature de la voyelle V_2 (/ i, e, ε, a, œ, ɔ /). Les tests post-hoc ont été effectués après les analyses de Mesures répétées, lorsque les résultats étaient significatifs.

7.3.1 Hypothèse de planification de type voisinage avec exécution séquentielle.

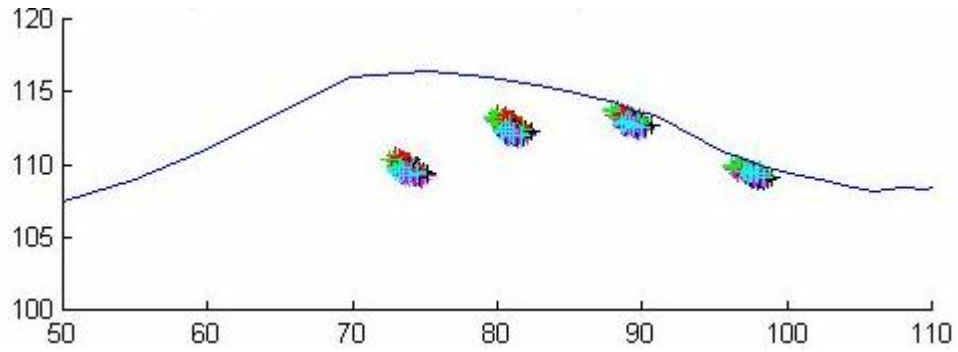
a. Effet de la voyelle V_2 sur la consonne /k/

La figure 7.21 (a) montre les positions des noeuds sélectionnés pour la consonne /k/ dans toutes les séquences $/V_1kV_2/$ pour l'hypothèse de planification de type voisinage avec exécution séquentielle. Nous observons que la langue est bien en contact avec le palais dans la région vélaire. Les différentes couleurs de point représentent les différentes voyelles V_2 : le vert pour /i/ ; le rouge pour /e/ ; le violet pour /ɛ/ ; le bleu pour /a/ ; le noir pour /ɔ/ ; le cyan pour /œ/. Les données sont présentées pour toutes les voyelles V_1 confondues.

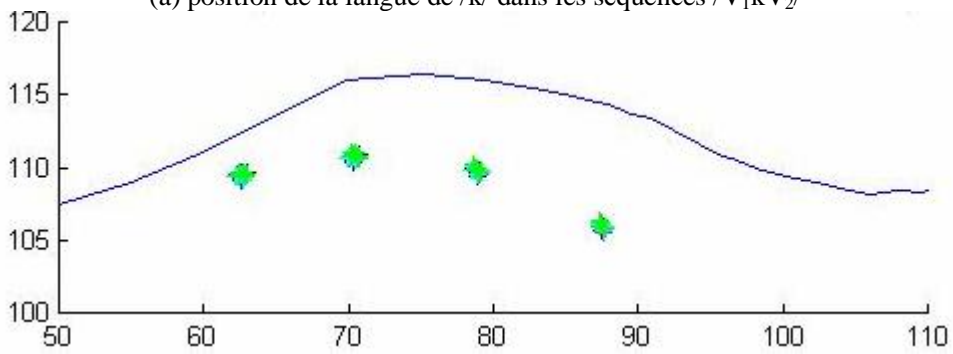
Les résultats d'analyses ANOVA sur ces données sont présentés dans le tableau 7.1 (première colonne). Nous observons que pour la consonne /k/, l'influence de V_2 est significative pour les 4 points du contour supérieur de langue. Le tableau 7.2 présente les résultats de tests post-hoc pour /k/. Il montre que la consonne /k/ est plus antérieure et plus haute pour $V_2=/i/$, $V_2=/e/$ $V_2=/ɛ/$ que pour $V_2=/a/$ et $V_2=/ɔ/$. Ceci est conforme à une stratégie d'anticipation.

b. Effet de la voyelle V_2 sur la voyelle V_1

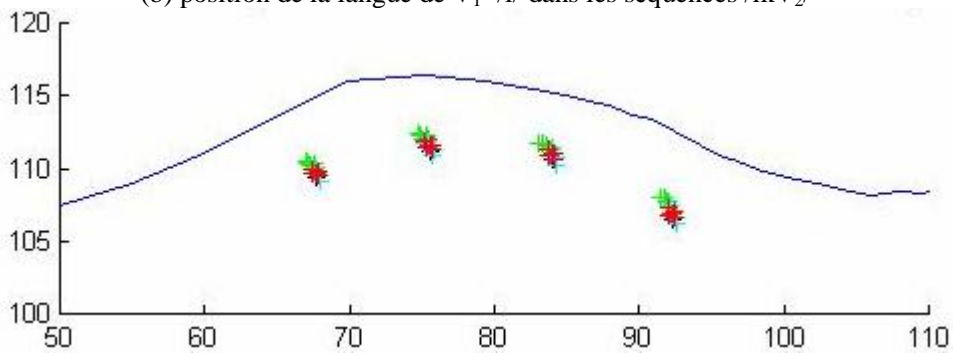
Les positions de la langue pour les six voyelles $V_1=/i, e, ɛ, a, œ, ɔ/$ sont présentées dans la figure 7.21 ((b), (c), (d), (e), (f), (g)) pour les différents contextes de V_2 (même code de couleur que pour figure 7.21 (a)). Les résultats d'analyses ANOVA sont présentés dans le tableau 7.1. Les tableaux 7.3, 7.4, 7.5, 7.6, 7.7, 7.8 montrent les résultats de tests post-hoc pour $V_1=/i, e, ɛ, a, œ, ɔ/$. Nous observons que la voyelle $V_1=/i/$ (en (b)) est moins influencée par la voyelle V_2 , mais elle est dans un sens cohérente avec l'idée d'anticipation. Pour la voyelle $V_1=/e/$ (en (c)), nous observons que la langue est plus antérieure et plus haute pour $V_2=/i/$ que pour les autres V_2 . La voyelle $V_1=/ɛ/$ (en (d)) montre une position plus postérieure et plus basse pour $V_2=/a/, /ɔ/$ que pour $V_2=/i/, /e/$. Ceci est aussi le cas pour la voyelle $V_1=/a/$ (en (e)), et $V_1=/œ/$ (en (f)). Pour la voyelle $V_1=/ɔ/$ (en (g)), l'effet de voyelle V_2 est moins fort. Dans la plupart des cas, les effets de la voyelle V_2 sur la voyelle V_1 sont cohérents avec une stratégie d'anticipation.



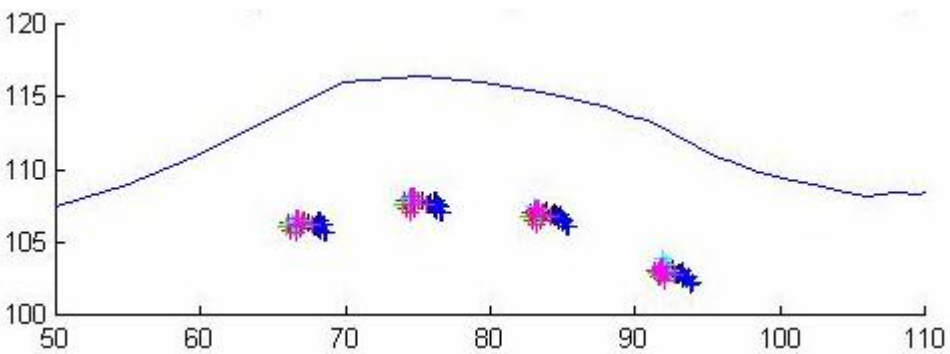
(a) position de la langue de /k/ dans les séquences $/V_1kV_2/$



(b) position de la langue de $V_1=/i/$ dans les séquences $/ikV_2/$



(c) position de la langue de $V_1=/e/$ dans les séquences $/ekV_2/$



(d) position de la langue de $V_1=/\epsilon/$ dans les séquences $/\epsilon kV_2/$

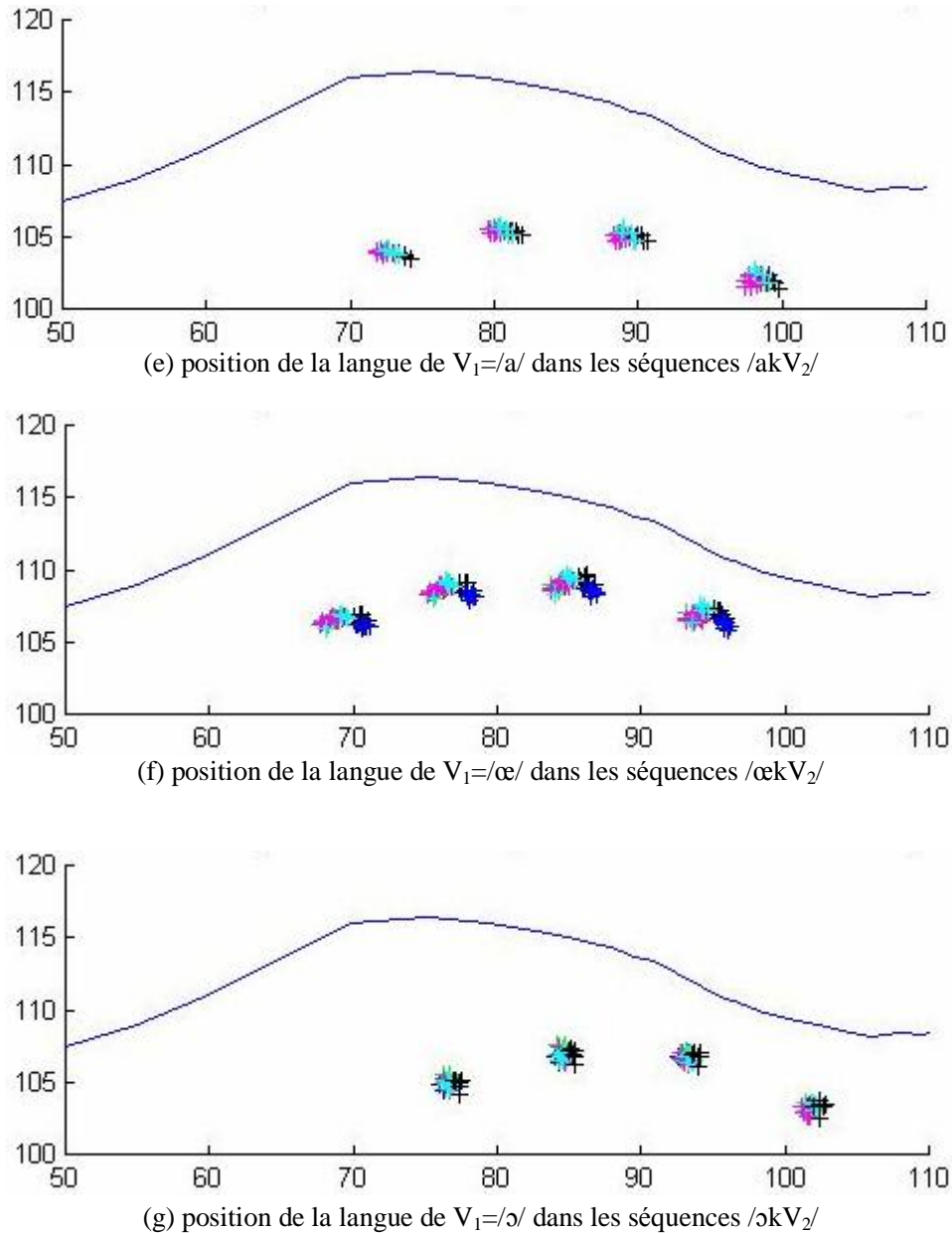


Figure 7.21 Simulations pour l'hypothèse de planification de type voisinage avec exécution séquentielle. Les positions des 4 nœuds du contour supérieur de la langue de la consonne /k/ et des voyelles V_1 / i, e, ε, a, œ, ɔ / sont présentées avec la forme du palais comme référence géométrique. Les différentes couleurs représentent le contexte V_2 (vert pour /i/, rouge /e/, violet pour /ε/, cyan pour /œ/, bleu pour /a/, noir pour /ɔ/). Lèvres à gauche –velum à droite.

Troisième partie : Résultats
Chapitre 7 Simulations de séquences VCV

Hypothèse voisinage	/k/	V ₁ =/i/	V ₁ =/e/	V ₁ =/ɛ/	V ₁ =/ a /	V ₁ =/ œ /	V ₁ =/ ɔ /
T1x	F=13.669 p=0		F=3.730 p=0.007	F=77.422 p=0	F=53.195 p=0	F=85.619 p=0	F=36.741 p=0
T1y	F=40.953 p=0	F=4.895 p=0.001	F=8.20 p=0		F=10.261 p=0	F=10.630 p=0	F=5.995 p=0
T2x	F=15.051 p=0		F=4.107 p=0.004	F=82.215 p=0	F=33.198 p=0	F=86.414 p=0	F=31.727 p=0
T2y	F=52.634 p=0	F=5.768 p=0	F=7.452 p=0	F=3.480 p=0.01	F=5.944 p=0	F=11.071 p=0	F=5.405 p=0.001
T3x	F=13.195 p=0		F=5.375 p=0.001	F=89.117 p=0	F=33.712 p=0	F=83.241 p=0	F=23.079 p=0
T3y	F=56.955 p=0	F=6.507 p=0	F=5.953 p=0	F=10.443 p=0	F=6.283 p=0	F=12.105 p=0	F=3.181 p=0
T4x	F=10.311 p=0	F=2.998 p=0.02	F=3.640 p=0.008	F=94.412 p=0	F=33.709 p=0	F=87.823 p=0	F=19.499 p=0
T4y	F=45.645 p=0	F=4.706 p=0.002	F=10.475 p=0	F=22.097 p=0	F=9.936 p=0	F=18.469 p=0	F=4.259 p=0.003

Tableau 7.1 Résultats d'analyses ANOVA pour l'hypothèse de planification de voisinage.
Seuls les cas significatifs sont présentés (p<0.05).

/k/ (en mm) Hypothèse de voisinage	V2 /ɔ/-/i/	/ɔ/-/e/	/ɔ/-/ɛ/	/ɔ/-/œ/	/a/-/i/	/a/-/e/	/a/-/ɛ/	/a/-/œ/	/œ/-/i/
T1x	0.440	0.365	0.204	0.291	0.442	0.368	0.206	0.293	0.149
T2x	0.482	0.329	0.272	0.323	0.470	0.366	0.251	0.361	0.109
T3x	0.458	0.302	0.313	0.356	0.381	0.225	0.236	0.279	0.102
T4x	0.399	0.339	0.231	0.220	0.367	0.306	0.198	0.187	0.179
	V2 /i/-/ɛ/	/i/-/a/	/i/-/œ/	/e/-/ɛ/	/e/-/a/	/e/-/œ/	/e/-/ɔ/	/ɔ/-/a/	/ɔ/-/œ/
T1y	0.233	0.257	0.196	0.422	0.446	0.384	0.251	0.195	0.133
T2y	0.279	0.256	0.254	0.515	0.493	0.490	0.268	0.225	0.222
T3y	0.347	0.278	0.308	0.445	0.376	0.407	0.240	0.136	0.166
T4y	0.239	0.303	0.120	0.207	0.272	0.288	0.201		

Tableau 7.2 Résultats de tests post-hoc pour /k/ pour l'hypothèse de planification de voisinage avec exécution séquentielle. Différence moyenne des positions de la langue (en mm) de /k/ dans les séquences /V₁kV₂/. La valeur est positive quand la position de langue dans le premier contexte V₂ est postérieure à ou plus haute que celles qui sont obtenues dans le deuxième contexte V₂. Seuls les cas statistiquement significatifs sont présentés (p<0.05)

V ₁ =/i/ (en mm) Hypothèse de voisinage	V2 /ɔ/-/i/	/a/-/i/	
T4x	0.156	0.152	
	V2 /i/-/a/	/e/-/a/	/e/-/ɔ/
T1y	0.215	0.214	
T2y	0.238	0.247	0.165
T3y	0.247	0.245	0.166
T4y	0.275	0.285	0.169

Tableau 7.3 Résultats de tests post-hoc pour V₁=/i/ pour l'hypothèse de planification de voisinage avec exécution séquentielle. (Voir Tableau 7.2 pour les détails) (en mm)

Troisième partie : Résultats
Chapitre 7 Simulations de séquences VCV

V ₁ =/e/ (en mm) Hypothèse de voisinage	V2 /ɔ/-/i/	/a/-/i/	/ɛ/-/i/	/e/-/i/	/œ/-/i/
T1x	0.239	0.254	0.294	0.284	0.299
T2x	0.264	0.271	0.323	0.201	0.305
T3x	0.317	0.324	0.376	0.251	0.336
T4x	0.250	0.249	0.301	0.291	0.306
	V2 /i/-/e/	/i/-/ɛ/	/i/-/a/	/i/-/œ/	/i/-/ɔ/
T1y	0.450	0.522	0.434	0.490	0.502
T2y	0.484	0.558	0.459	0.512	0.543
T3y	0.502	0.574	0.464	0.499	0.567
T4y	0.703	0.790	0.683	0.712	0.774

Tableau 7.4 Résultats de tests post-hoc pour V₁=/e/ pour l'hypothèse de planification de voisinage avec exécution séquentielle. (Voir Tableau 7.2 pour les détails) (en mm)

V ₁ =/ ε / (en mm) Hypothèse de voisinage	V2 /ɔ/-/i/	/ɔ/-/e/	/ɔ/-/ɛ/	/ɔ/-/œ/	/a/-/i/	/a/-/e/	/a/-/ɛ/	/a/-/œ/	/a/-/ɔ/
T1x	1.425	1.295	1.150	1.313	1.909	1.779	1.634	1.797	0.484
T2x	1.453	1.326	1.187	1.315	1.987	1.860	1.721	1.848	0.543
T3x	1.391	1.276	1.162	1.212	1.917	1.802	1.689	1.738	0.526
T4x	1.271	1.157	1.071	0.950	1.766	1.651	1.565	1.444	0.495
	V2 /i/-/a/	/e/-/a/	/ɛ/-/a/	/œ/-/a/	/ɔ/-/a/				
T2y	0.302	0.300	0.361	0.334	0.322				
T3y	0.310	0.333	0.354	0.362	0.308				
T4y	0.304	0.307	0.363	0.617	0.265				

Tableau 7.5 Résultats de tests post-hoc pour V₁=/ ε / pour l'hypothèse de planification de voisinage avec exécution séquentielle. (Voir Tableau 7.2 pour les détails) (en mm)

V ₁ =/a/ (en mm) Hypothèse de voisinage	V2 /ɔ/-/ɛ/	/ɔ/-/a/	/ɔ/-/œ/	/a/-/ɛ/	/œ/-/ɛ/
T1x	1.495	0.736	0.776	0.759	0.720
T2x	1.469	0.759	0.772	0.711	0.679
T3x	1.481	0.799	0.818	0.682	0.663
T4x	1.492	0.773	0.725	0.719	0.760
	V2 /i/-/ɛ/	/i/-/ɔ/	/e/-/ɔ/	/œ/-/ɔ/	
T1y	0.193	0.364	0.145	0.309	
T2y	0.244	0.300	0.146	0.315	
T3y	0.281	0.281	0.149	0.304	
T4y	0.302	0.429	0.265	0.475	

Tableau 7.6 Résultats de tests post-hoc pour V₁=/ a / pour l'hypothèse de planification de voisinage avec exécution séquentielle. (Voir Tableau 7.2 pour les détails) (en mm)

Troisième partie : Résultats
Chapitre 7 Simulations de séquences VCV

V ₁ =/œ/ Hypothèse voisinage	V2 /ɔ/-i/	/ɔ/-e/	/ɔ/-ɛ/	/ɔ/-œ/	/a/-i/	/a/-e/	/a/-ɛ/	/a/-œ/	/a/-ɔ/	/œ/-i/	/œ/-ɛ/
T1x	2.161	1.588	2.067	1.221	2.473	1.899	2.378	1.532	0.312	0.940	0.846
T2x	2.048	1.506	1.942	1.202	2.443	1.902	2.337	1.597	0.395	0.846	0.740
T3x	1.967	1.437	1.857	1.170	2.404	1.874	2.294	1.607	0.437	0.797	0.687
T4x	1.834	1.340	1.764	1.055	2.278	1.784	2.208	1.498	0.444	0.779	0.709
	V2 /i/-a/	/e/-a/	/ɛ/-a/	/œ/-a/	/ɔ/-a/						
T1y	0.428	0.412	0.368	0.583	0.489						
T2y	0.387	0.404	0.426	0.520	0.522						
T3y	0.420	0.439	0.429	0.537	0.524						
T4y	0.559	0.566	0.464	0.569	0.549						

Tableau 7.7 Résultats de tests post-hoc pour V₁=/œ/ pour l'hypothèse de planification de voisinage avec exécution séquentielle. (Voir Tableau 7.2 pour les détails) (en mm)

V ₁ =/ɔ/ (en mm) Hypothèse de voisinage	V2 /ɔ/-i/	/ɔ/-e/	/ɔ/-ɛ/	/ɔ/-œ/	/œ/-i/	/œ/-ɛ/
T1x	0.769	0.684	0.702	0.679		
T2x	0.792	0.776	0.737	0.686	0.206	
T3x	0.885	0.795	0.840	0.621	0.364	0.320
T4x	0.988	0.808	0.901	0.402	0.585	0.508
	V2 /i/-a/	/e/-a/	/œ/-a/	/ɔ/-a/		
T1y	0.495	0.582	0.369	0.384		
T2y	0.510	0.646	0.293	0.439		
T3y		0.323	0.429	0.213		
T4y		0.248	0.224	0.222		

Tableau 7.8 Résultats de tests post-hoc pour V₁=/ɔ/ pour l'hypothèse de planification de voisinage avec exécution séquentielle. (Voir Tableau 7.2 pour les détails) (en mm)

7.3.2 Hypothèse de planification syllabique avec exécution séquentielle.

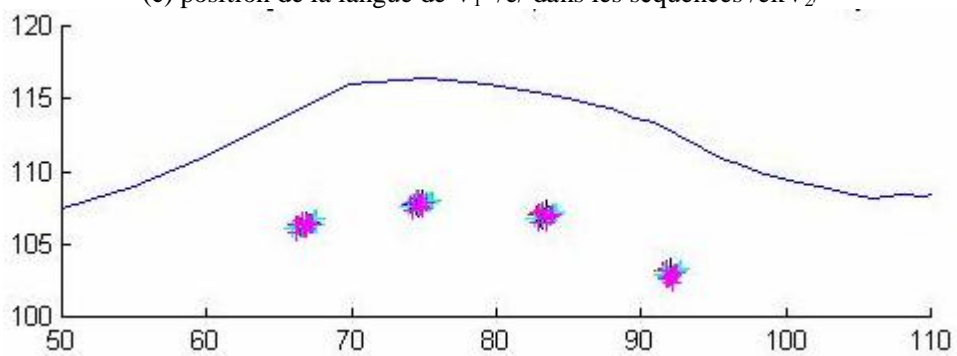
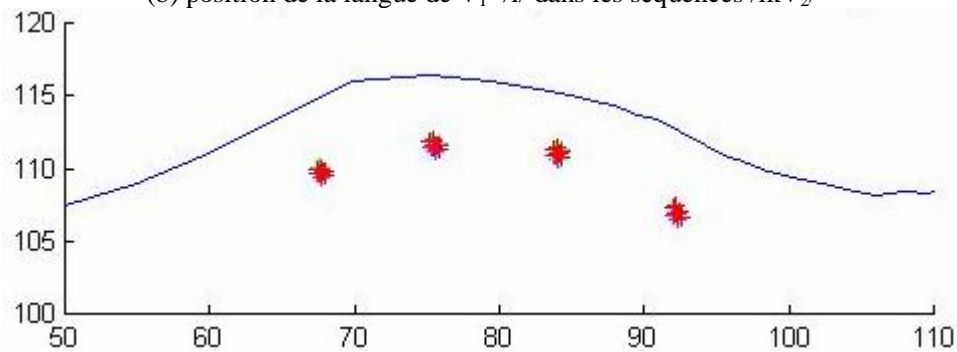
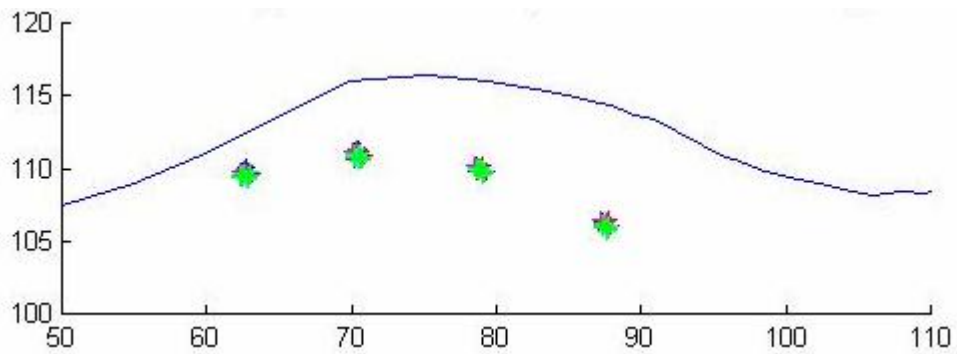
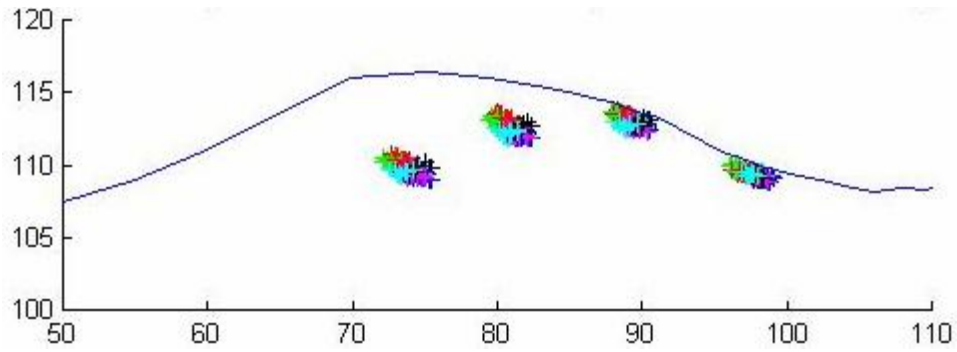
a. Effet de la voyelle V_2 sur la consonne /k/

La figure 7.22 (a) montre les positions des 4 nœuds de la langue pour la consonne /k/ dans toutes les séquences $/V_1kV_2/$ pour l'hypothèse de planification syllabique. Les résultats d'analyses ANOVA sont présentés dans le tableau 7.9. Le tableau 7.10 présente des résultats de tests post-hoc. Il montre que la consonne /k/ est plus antérieure et plus haute pour $V_2=/i/$, $V_2=/e/$ $V_2=/\epsilon/$ que pour $V_2=/a/$ et $V_2=/\text{ɔ}/$. Nous constatons que la consonne /k/ est clairement plus influencée par V_2 pour cette hypothèse syllabique que pour l'hypothèse du voisinage (tableau 7.1). Les différences moyennes de position de la langue sont systématiquement plus grandes pour l'hypothèse syllabique que pour l'hypothèse du voisinage (tableau 7.2). Cette différence est particulièrement nette sur la direction horizontale.

b. Effet de la voyelle V_2 sur la voyelle V_1

Les positions des 4 nœuds de la langue pour les six voyelles $V_1=/i, e, \epsilon, a, \text{œ}, \text{ɔ}/$ sont présentés dans la figure 7.22 ((b), (c), (d), (e), (f), (g)). Nous observons que pour l'hypothèse syllabique, aucune voyelle V_1 n'est significativement influencée par la voyelle V_2 .

Troisième partie : Résultats
Chapitre 7 Simulations de séquences VCV



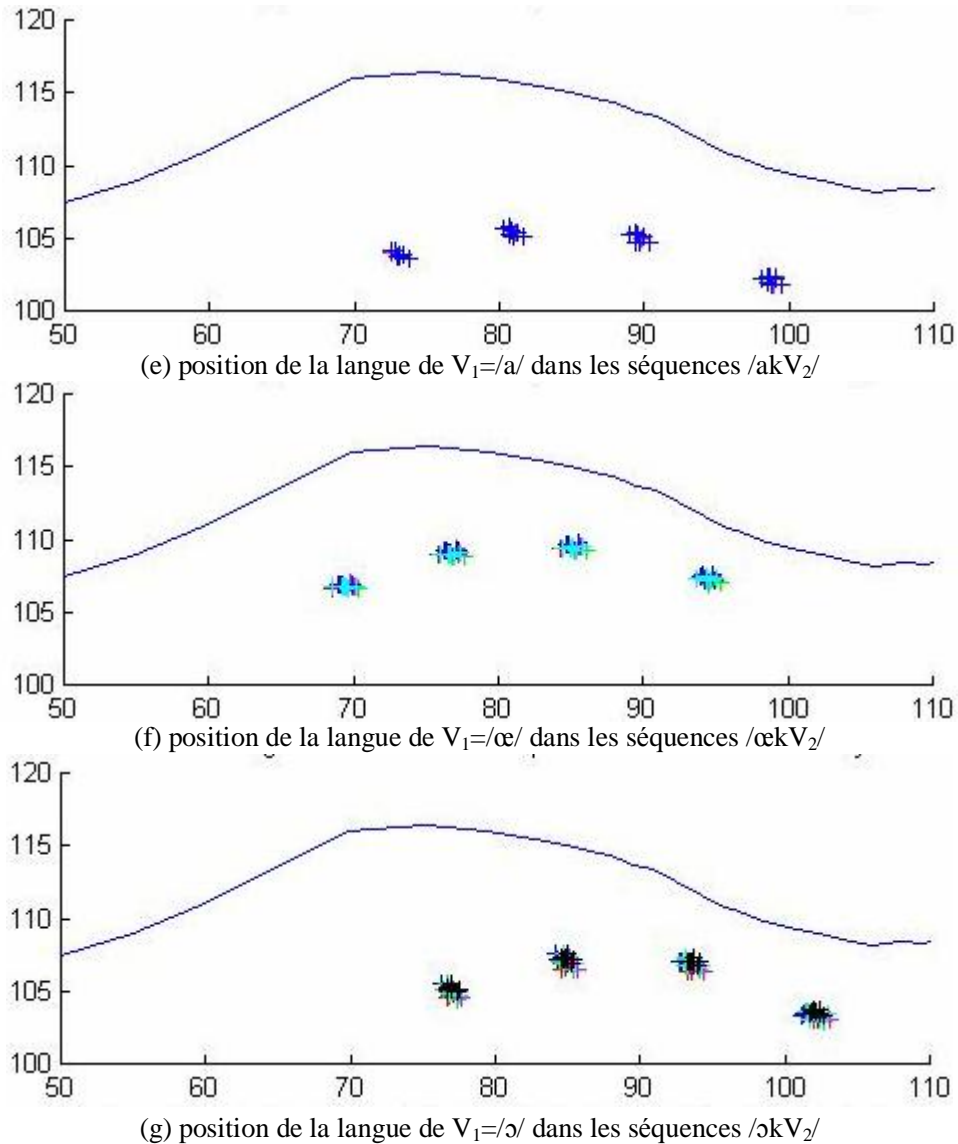


Figure 7.22 Simulations pour l'hypothèse de planification syllabique avec exécution séquentielle. Les positions des 4 nœuds du contour supérieur de la langue de la consonne /k/ et des voyelles V_1 / i, e, ɛ, a, α, ɔ / sont présentées avec la forme du palais comme référence géométrique. Les différentes couleurs représentent le contexte V_2 (vert pour /i/, rouge /e/, violet pour /ɛ/, cyan pour /æ/, bleu pour /a/, noir pour /ɔ/). Lèvres à gauche –velum à droite.

Troisième partie : Résultats
Chapitre 7 Simulations de séquences VCV

Hypothèse syllabique	/k/
T1x	F=67.345 p=0
T1y	F=79.694 p=0
T2x	F=65.285 p=0
T2y	F=122.491 p=0
T3x	F=69.245 p=0
T3y	F=131.068 p=0
T4x	F=71.326 p=0
T4y	F=119.818 p=0

Tableau 7.9 Résultats d'analyses ANOVA pour l'hypothèse de planification syllabique avec exécution séquentielle. (Voir Tableau 7.1 pour les détails)

/k/ (en mm) Hypothèse syllabique	V2 /ɔ/-/i/	/ɔ/-/e/	/ɔ/-/ɛ/	/ɔ/-/œ/	/a/-/i/	/a/-/e/	/a/-/ɛ/	/a/-/œ/	/œ/-/i/	/œ/-/e/
T1x	0.957	0.895	0.451	0.452	1.012	0.950	0.506	0.507	0.505	0.442
T2x	0.984	0.815	0.539	0.616	0.966	0.797	0.522	0.598	0.369	0.199
T3x	1.021	0.829	0.507	0.584	1.028	0.836	0.513	0.591	0.437	0.245
T4x	1.024	0.870	0.433	0.547	1.110	0.956	0.519	0.633	0.477	0.323
	V2 /i/-/ɛ/	/i/-/a/	/i/-/œ/	/i/-/ɔ/	/e/-/ɛ/	/e/-/a/	/e/-/œ/	/e/-/ɔ/	/ɔ/-/a/	/ɔ/-/œ/
T1y	0.314	0.406	0.189	0.373	0.238	0.330	0.113	0.297		
T2y	0.476	0.489	0.474	0.363	0.588	0.602	0.586	0.448	0.153	0.138
T3y	0.461	0.539	0.482	0.255	0.770	0.848	0.791	0.564	0.283	0.227
T4y	0.410	0.530	0.414	0.273	0.663	0.784	0.667	0.527	0.257	0.140

Tableau 7.10 Résultats de tests post-hoc pour /k/ pour l'hypothèse de planification syllabique avec exécution séquentielle. (Voir Tableau 7.2 pour les détails) (en mm)

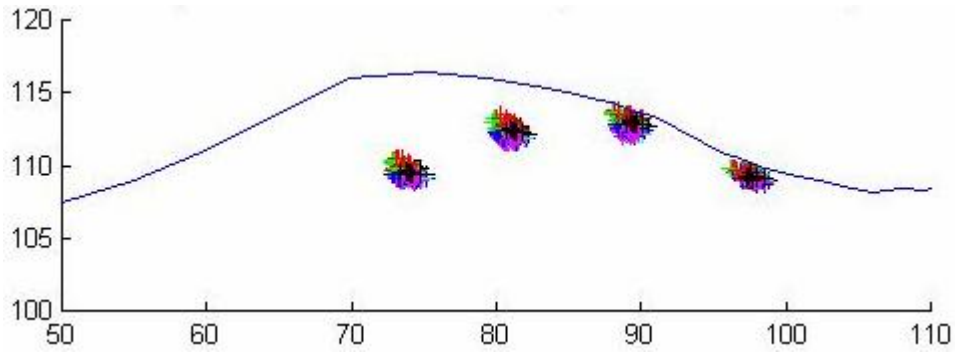
7.3.3 Hypothèse de planification inspirée du modèle d'Öhman avec exécution selon Öhman

a. Effet de la voyelle V_2 sur la consonne /k/

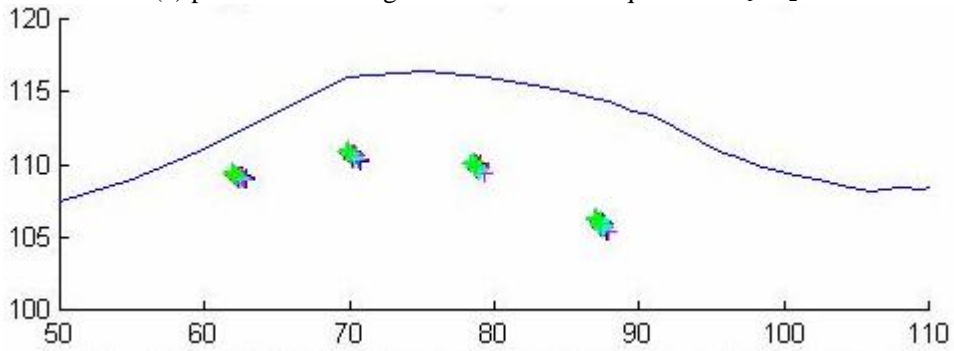
La figure 7.23 (a) montre les positions des 4 nœuds de la langue pour la consonne /k/ dans toutes les séquences $/V_1kV_2/$ pour l'hypothèse planification inspirée du modèle d'Öhman. Les résultats d'analyses ANOVA sont présentés dans le tableau 7.11 (première colonne). Le tableau 7.12 présente des résultats de tests post-hoc. Nous trouvons que la consonne /k/ est plus antérieure et plus haute pour $V_2=/i/$, $V_2=/e/$ $V_2=/\epsilon/$ que pour $V_2=/a/$ et $V_2=/\text{o}/$. Cependant, nous observons que la position de /k/ est un peu moins systématiquement influencée par V_2 pour l'hypothèse inspirée du modèle d'Öhman que pour l'hypothèse de voisinage. Cette différence est notamment sur la direction horizontale qui est une direction dominante de l'influence anticipatoire pour la consonne vélaire /k/ (voir l'observation sur les sujets dans le chapitre 6, et une discussion dans le chapitre 8).

b. Effet de la voyelle V_2 sur la voyelle V_1

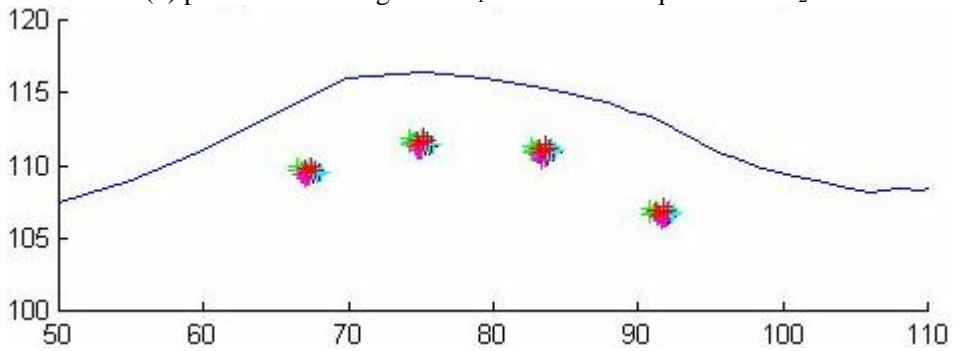
Les positions de 4 nœuds de la langue pour les six voyelles $V_1=/i, e, \epsilon, a, \text{œ}, \text{o}/$ sont présentées dans la figure 7.23 ((b), (c), (d), (e), (f), (g)). Les résultats d'analyses ANOVA sont présentés dans le tableau 7.11. Les tableaux 7.13, 7.14, 7.15, 7.16, 7.17, 7.18 montrent les résultats de tests post-hoc pour $V_1=/i, e, \epsilon, a, \text{œ}, \text{o}/$. Nous observons une plus grande influence de la voyelle V_2 sur la voyelle V_1 . Si nous comparons l'hypothèse de voisinage (figure 7.21), les voyelles $V_1=/i/$, $V_1=/e/$, $V_1=/a/$, $V_1=/\text{œ}/$ et $V_1=/\text{o}/$ montrent plus de variabilité en fonction de V_2 .



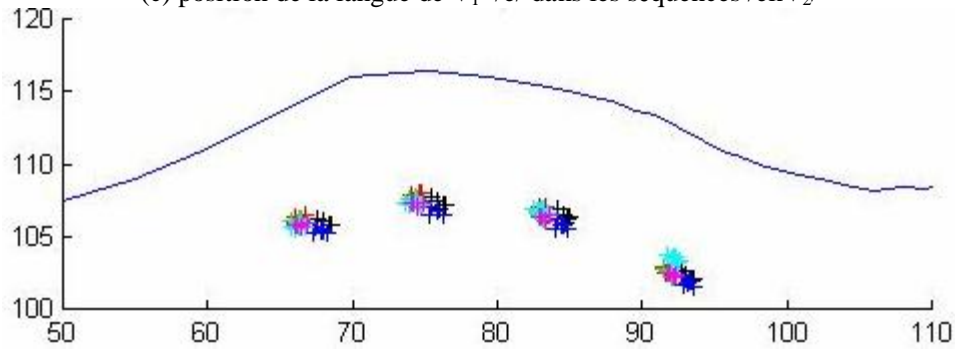
(a) position de la langue de /k/ dans les séquences $/V_1kV_2/$



(b) position de la langue de $V_1=/i/$ dans les séquences $/ikV_2/$



(c) position de la langue de $V_1=/e/$ dans les séquences $/ekV_2/$



(d) position de la langue de $V_1=/\epsilon/$ dans les séquences $/\epsilon kV_2/$

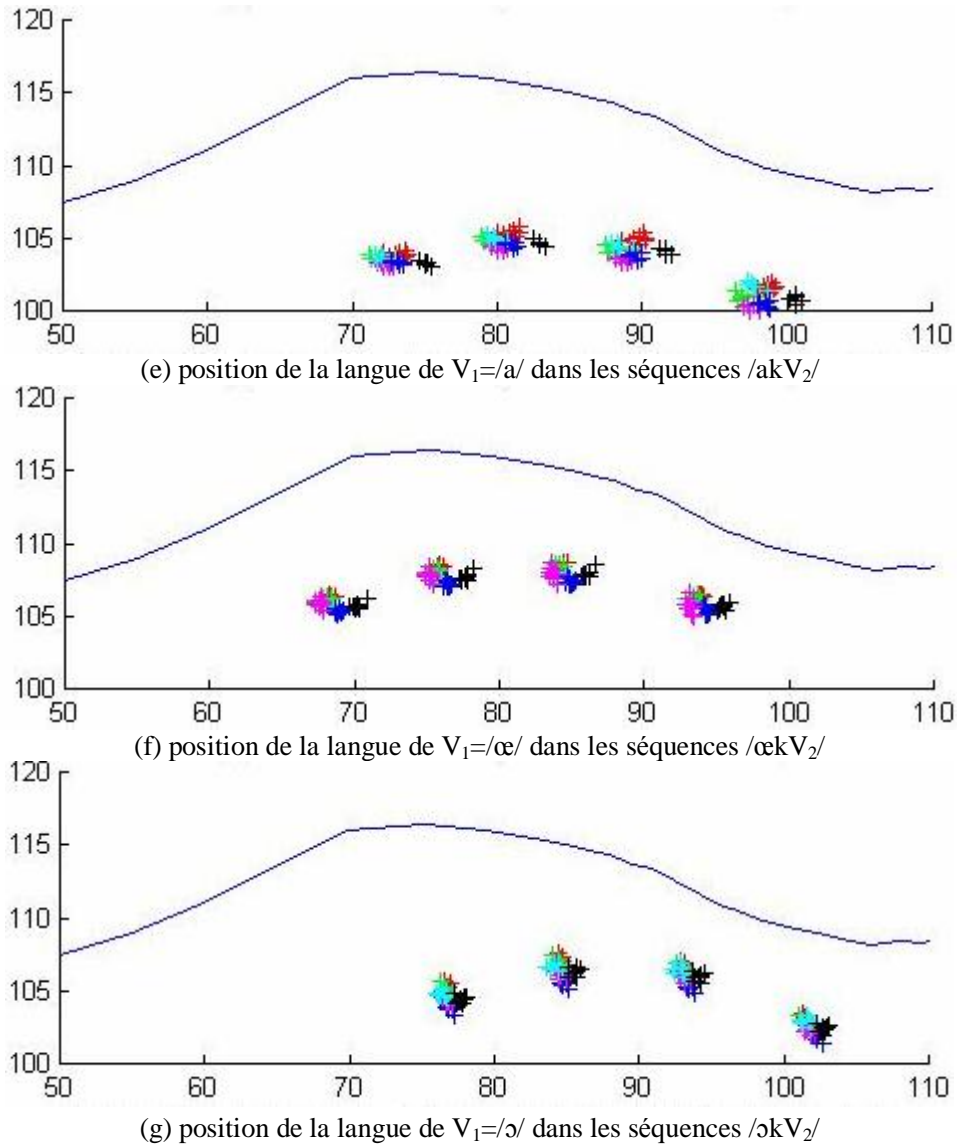


Figure 7.23 Simulations pour l'hypothèse de planification inspirée du modèle d'Öhman avec exécution selon Öhman. Les positions des 4 nœuds du contour supérieur de la langue de la consonne /k/ et des voyelles V_1 /i, e, ε, a, œ, ɔ/ sont présentées avec la forme du palais comme référence géométrique. Les différentes couleurs représentent le contexte V_2 (vert pour /i/, rouge /e/, violet pour /ε/, cyan pour /œ/, bleu pour /a/, noir pour /ɔ/). Lèvres à gauche –velum à droite.

Troisième partie : Résultats
Chapitre 7 Simulations de séquences VCV

Hypothèse selon Öhman	/k/	V ₁ =/i/	V ₁ =/e/	V ₁ =/ε/	V ₁ =/ a /	V ₁ =/ œ /	V ₁ =/ ɔ /
T1x	F=6.419 p=0	F=5.073 p=0.001	F=25.000 p=0	F=85.487 p=0	F=79.432 p=0	F=151.524 p=0	F=22.137 p=0
T1y	F=39.020 p=0		F=10.617 p=0	F=22.708 p=0	F=32.41 p=0	F=36.425 p=0	F=43.856 p=0
T2x	F=9.997 p=0	F=4.498 p=0.002	F=24.315 p=0	F=91.168 p=0	F=76.631 p=0	F=155.847 p=0	F=21.125 p=0
T2y	F=55.37 p=0		F=11.666 p=0	F=30.765 p=0	F=44.923 p=0	F=33.947 p=0	F=35.576 p=0
T3x	F=12.281 p=0	F=4.650 p=0.002	F=24.356 p=0	F=95.936 p=0	F=76.937 p=0	F=150.052 p=0	F=17.371 p=0
T3y	F=60.4 p=0		F=12.922 p=0	F=38.854 p=0	F=57.924 p=0	F=28.181 p=0	F=31.309 p=0
T4x	F=9.050 p=0	F=6.544 p=0	F=25.674 p=0	F=113.229 p=0	F=78.305 p=0	F=163.557 p=0	F=16.928 p=0
T4y	F=37.59 p=0		F=13.769 p=0	F=93.208 p=0	F=80.670 p=0	F=21.781 p=0	F=26.325 p=0

Tableau 7.11 Résultats d’analyses ANOVA pour l’hypothèse de planification inspirée du modèle d’Öhman avec exécution selon Öhman. (Voir Tableau 7.1 pour les détails)

/k/ (en mm) Hypothèse selon Öhman	V2 /ɔ/-/i/	/ɔ/-/e/	/ɔ/-/ε/	/ɔ/-/œ/	/a/-/i/	/a/-/e/	/œ/-/i/		
T1x	0.319	0.274			0.279	0.233	0.222		
T2x	0.474	0.259	0.268	0.296	0.292		0.178		
T3x	0.502	0.297	0.219	0.182	0.364	0.160	0.320		
T4x	0.452	0.326	0.140	0.189	0.382	0.256	0.263		
	V2 /i/-/ε/	/i/-/a/	/i/-/œ/	/e/-/ε /	/e/-/a/	/e/-/œ/	/e/-/ɔ/	/ɔ/-/a/	/ɔ/-/œ/
T1y	0.345	0.231	0.268	0.288	0.274	0.211	0.226		
T2y	0.476	0.323	0.448	0.536	0.483	0.507	0.251	0.332	0.457
T3y	0.347	0.260	0.331	0.652	0.565	0.637	0.259	0.306	0.377
T4y	0.292	0.246	0.204	0.553	0.507	0.465	0.253	0.255	0.212

Tableau 7.12 Résultats de tests post-hoc pour /k/ pour l’hypothèse de planification inspirée du modèle d’Öhman avec exécution selon Öhman. (Voir Tableau 7.2 pour les détails) (en mm)

V ₁ =/i/ (en mm) Hypothèse selon Öhman	V2 /ɔ/-/i/	/œ/-/i/	/a/-/i/	/ε/-/i/	/e/-/i/
T1x	0.449	0.274	0.475	0.268	0.308
T2x	0.445	0.272	0.471	0.266	0.385
T3x	0.416	0.250	0.437	0.261	0.360
T4x	0.390	0.246	0.418	0.263	0.329

Tableau 7.13 Résultats de tests post-hoc pour V₁=/i/ pour l’hypothèse de planification inspirée du modèle d’Öhman avec exécution selon Öhman. (Voir Tableau 7.2 pour les détails) (en mm)

Troisième partie : Résultats
Chapitre 7 Simulations de séquences VCV

V ₁ =/e/ (en mm) Hypothèse selon Öhman	V2 /ɔ/-/i/	/ɔ/-/ε/	/œ/-/i/	/œ/-/ ε /	/a/-/i/	/a/-/ε/	/ε/-/i/	/ε/-/i/
T1x	0.847	0.343	1.002	0.498	0.923	0.419	0.504	0.743
T2x	0.849	0.317	1.014	0.482	0.930	0.398	0.523	0.740
T3x	0.833	0.311	1.006	0.484	0.909	0.386	0.523	0.729
T4x	0.801		0.949	0.287	0.890		0.564	0.701
	V2 /i/-/ε/	/i/-/a/	/e/-/ε/	/e/-/a/	/œ/-/ ε /	/œ/-/a/	/ɔ/-/ε/	
T1y	0.439	0.220	0.404	0.185	0.420	0.301	0.285	
T2y	0.474	0.204			0.522	0.252	0.350	
T3y	0.523	0.199	0.612	0.288	0.681	0.356		
T4y	0.529	0.182	0.574	0.227	0.715	0.368		

Tableau 7.14 Résultats de tests post-hoc pour V₁=/e/ pour l'hypothèse de planification inspirée du modèle d'Öhman avec exécution selon Öhman. (Voir Tableau 7.2 pour les détails) (en mm)

V ₁ =/ ε / (en mm) Hypothèse selon Öhman	V2 /ɔ/-/i/	/ɔ/-/e/	/ɔ/-/ε/	/ɔ/-/œ/	/a/-/i/	/a/-/e/	/a/-/ε/	/a/-/œ/	/ ε /-/i/
T1x	1.842	1.710	1.690	1.723	1.633	1.501	1.480	1.541	0.153
T2x	1.847	1.746	1.673	1.733	1.690	1.590	1.517	1.576	0.158
T3x	1.769	1.701	1.590	1.602	1.652	1.584	1.476	1.485	0.179
T4x	1.588	1.599	1.365	1.061	1.578	1.590	1.356	1.051	0.223
	V2 /i/-/ε/	/i/-/a/	/i/-/ɔ/	/e/-/ε /	/e/-/a/	/e/-/œ/	/e/-/ɔ/	/ɔ/-/a/	/œ/-/a/
T1y	0.206	0.577	0.277	0.415	0.786	0.476	0.386	0.400	0.310
T2y	0.234	0.768	0.288	0.453	0.987	0.294	0.507	0.480	0.692
T3y	0.257	0.858	0.346	0.474	1.074		0.562	0.512	1.091
T4y	0.168	0.731	0.369	0.286	0.850		0.488	0.362	1.685

Tableau 7.15 Résultats de tests post-hoc pour V₁=/ ε / pour l'hypothèse de planification inspirée du modèle d'Öhman avec exécution selon Öhman. (Voir Tableau 7.2 pour les détails) (en mm)

V ₁ =/a/ Hypothèse selon Öhman	V2 /ɔ/-/i/	/ɔ/-/e/	/ɔ/-/ε/	/ɔ/-/a/	/ɔ/-/œ/	/a/-/i/	/a/-/ ε /	/a/-/œ/	/e/-/i/	/ ε /-/i/	/œ/-/i/
T1x	2.985	1.699	2.498	1.705	2.213	1.280	0.793	0.507	1.286	0.486	0.772
T2x	2.934	1.666	2.240	1.668	2.271	1.266	0.752	0.603	1.268	0.514	0.664
T3x	3.015	1.664	2.507	1.726	2.356	1.289	0.781	0.630	0.843	0.508	0.659
T4x	3.121	1.655	2.683	1.800	2.080	1.321	0.883	0.280	1.027	0.438	1.041
	V2 /i/-/ε/	/i/-/a/	/e/-/ε/	/e/-/a/	/e/-/œ/	/e/-/ɔ/	/ɔ/-/a/	/œ/-/a/	/œ/-/ɔ/		
T1y	0.458	0.295	0.562	0.399	0.144	0.393		0.256	0.250		
T2y	0.526	0.215	0.937	0.590	0.330	0.342	0.248	0.259			
T3y	0.723	0.154	1.364	0.796	0.336	0.351	0.445	0.459			
T4y	0.885	0.185	1.646	0.946		0.575	0.371	1.034	0.663		

Tableau 7.16 Résultats de tests post-hoc pour V₁=/ a / pour l'hypothèse de planification inspirée du modèle d'Öhman avec exécution selon Öhman. (Voir Tableau 7.2 pour les détails) (en mm)

Troisième partie : Résultats
Chapitre 7 Simulations de séquences VCV

V ₁ =/oe/ (mm) Hypothèse selon Öhman	V2 /ɔ/-/i/	/ɔ/-/e/	/ɔ/-/ɛ/	/ɔ/-/a/	/ɔ/-/œ/	/a/-/i/	/a/-/e/	/a/-/ɛ/	/a/-/œ/	/œ/-/i/	/œ/-/ɛ/
T1x	2.558	1.951	2.709	1.233	2.025	1.326	0.628	1.476	0.792		
T2x	2.517	1.909	2.602	1.170	2.026	1.347	0.739	1.423	0.856	0.491	0.576
T3x	2.456	1.870	2.516	1.139	2.023	1.327	0.723	1.377	0.884	0.443	0.493
T4x	2.360	1.791	2.409	1.049	1.886	1.311	0.742	1.360	0.837	0.474	0.523
	V2 /i/-/ɛ/	/i/-/a/	/i/-/œ/	/i/-/ɔ/	/e/-/a/	/ɛ/-/a/	/œ/-/a/	/ɔ/-/a/			
T1y	0.257	0.756	0.319	0.377	0.997	0.499	0.437	0.380			
T2y	0.335	0.929	0.352	0.450	1.212	0.594	0.577	0.479			
T3y	0.436	1.031	0.304	0.508	1.331	0.595	0.727	0.523			
T4y	0.514	0.976		0.627	1.189	0.462	0.425	0.349			

Tableau 7.17 Résultats de tests post-hoc pour V₁=/œ / pour l'hypothèse de planification inspirée du modèle d'Öhman avec exécution selon Öhman. (Voir Tableau 7.2 pour les détails) (en mm)

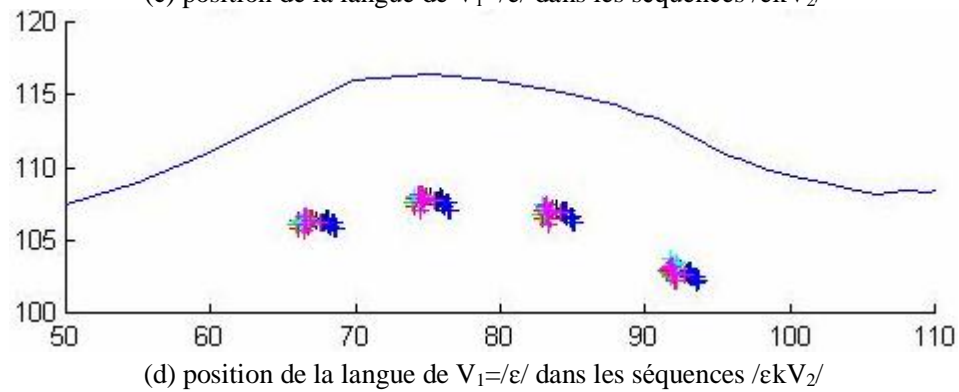
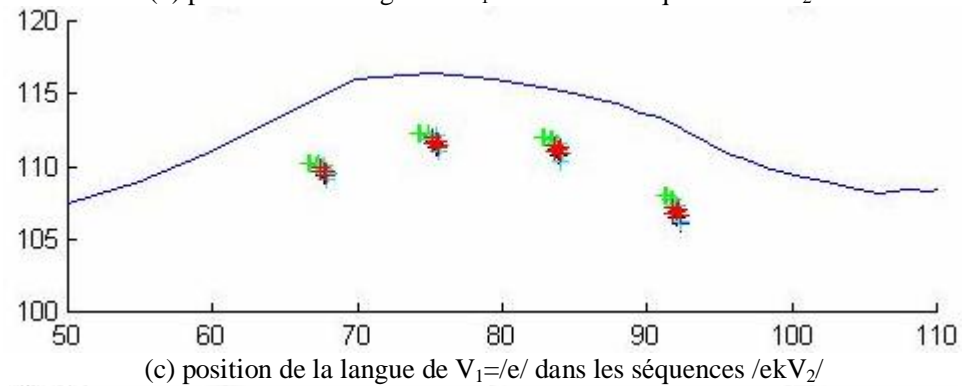
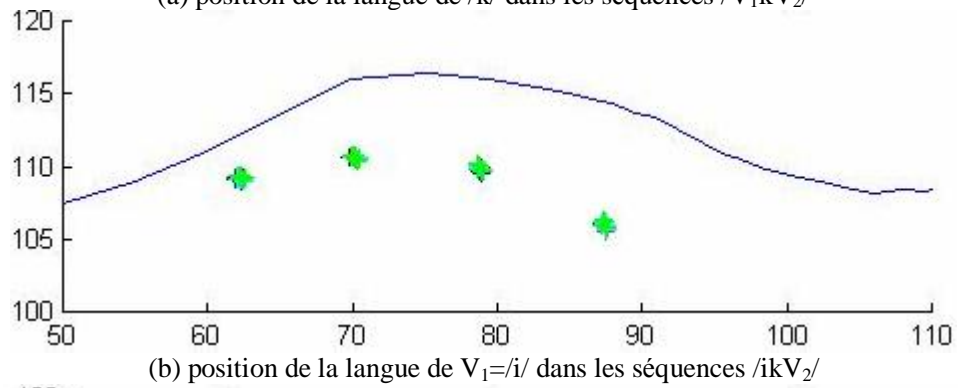
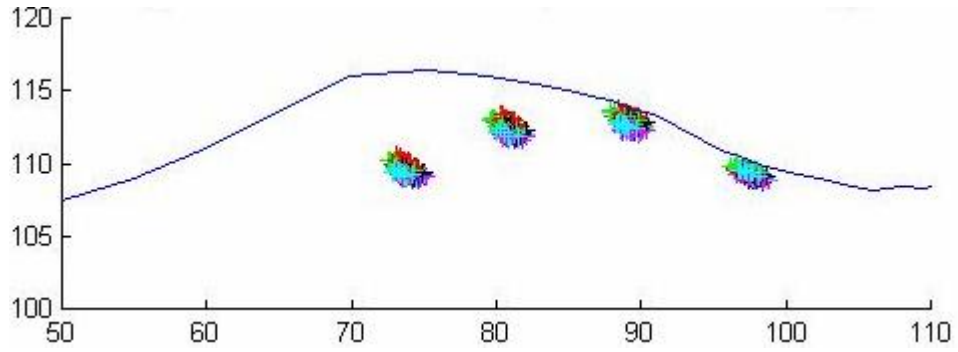
V ₁ =/ɔ/ (mm) Hypothèse selon Öhman	V2 /ɔ/-/i/	/ɔ/-/e/	/ɔ/-/ ɛ /	/ɔ/-/a/	/ɔ/-/œ/			
T1x	0.997	0.817	0.906	0.902	0.937			
T2x	1.043	0.876	0.949	1.020	0.973			
T3x	1.057	0.952	0.993	0.938	0.856			
T4x	1.042	1.027	1.000	0.648	0.654			
	V2 /i/-/ɛ/	/i/-/a/	/i/-/ɔ/	/e/-/a/	/ ɛ/-/a/	/œ/-/a/	/ɔ/-/a/	/œ/-/ɔ/
T1y	0.343	0.885	0.406	1.005	0.542	0.844	0.479	0.365
T2y	0.384	1.020	0.363	1.186	0.636	0.878	0.657	0.221
T3y	0.290	0.950	0.280	1.088	0.660	0.995	0.670	0.324
T4y	0.238	0.629	0.237	0.897	0.391	0.897	0.392	0.505

Tableau 7.18 Résultats de tests post-hoc pour V₁=/ ɔ / pour l'hypothèse de planification inspirée du modèle d'Öhman avec exécution selon Öhman. (Voir Tableau 7.2 pour les détails) (en mm)

7.3.4 Deux autres modèles s'appuyant sur des hypothèses de planification de voisinage avec exécution selon Öhman

La figure 7.24 montre les simulations de l'hypothèse de planification de voisinage pour l'exécution selon Öhman avec 15ms de temps de départ vers la consonne. Nous n'avons observé aucune différence en comparant l'hypothèse de planification de voisinage avec l'exécution séquentielle (Figure 7.21).

La figure 7.25 montre les simulations pour l'exécution selon Öhman avec un essai de 50ms de temps de départ vers la consonne. Si nous comparons l'hypothèse de planification de voisinage avec l'exécution séquentielle (figure 7.21), nous observons que la consonne /k/ et les voyelles /i/, /e/, /a/, /æ/ sont plus affectées par la voyelle V₂. Mais comme nous l'avons dit plus haut, cet essai d'exécution n'est pas très réaliste. Il prouve que le temps de départ de la consonne sur la base vocalique d'exécution selon Öhman est important.



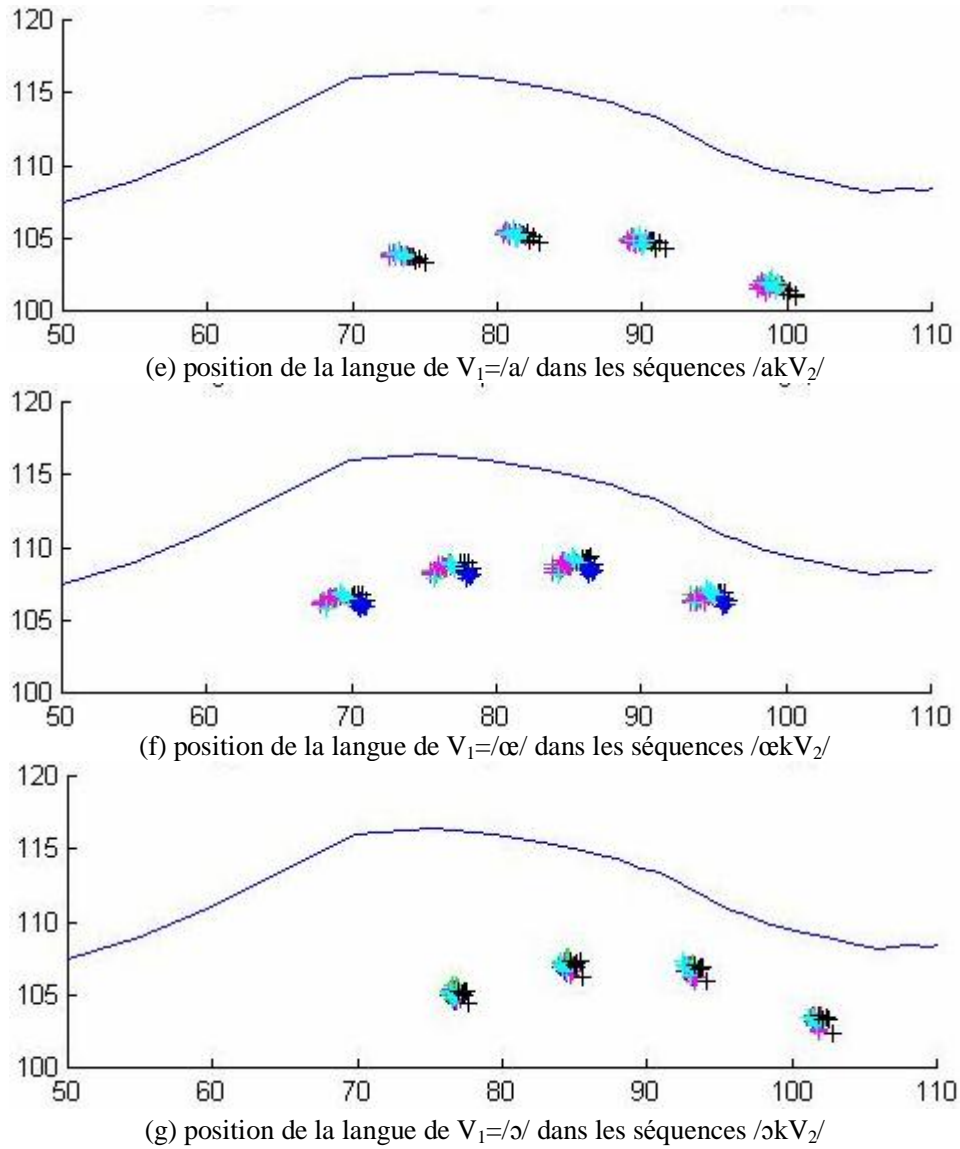
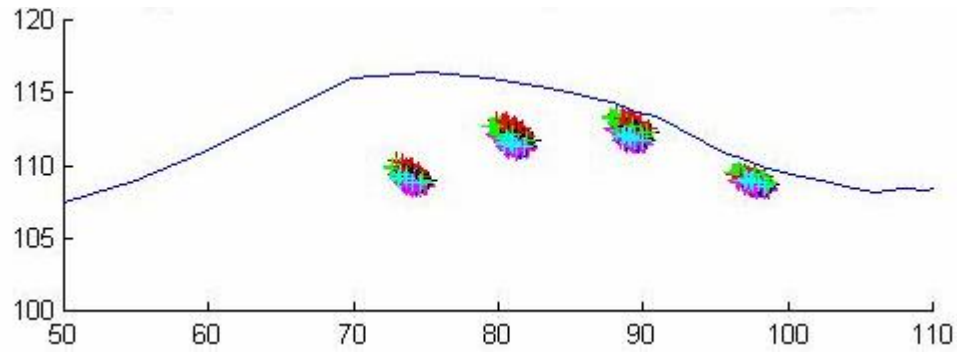
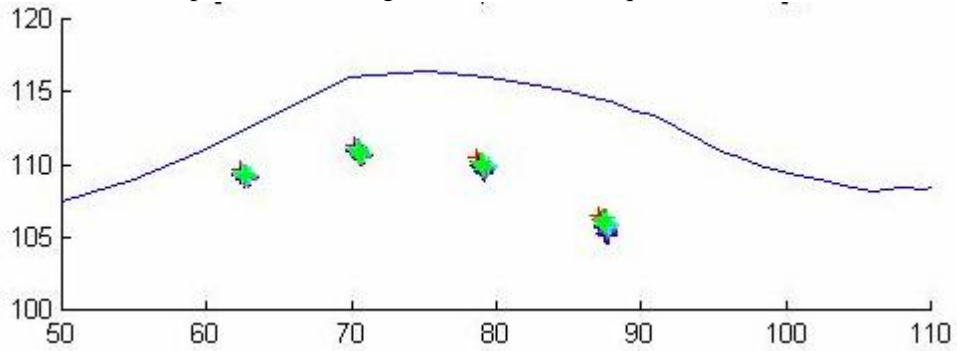


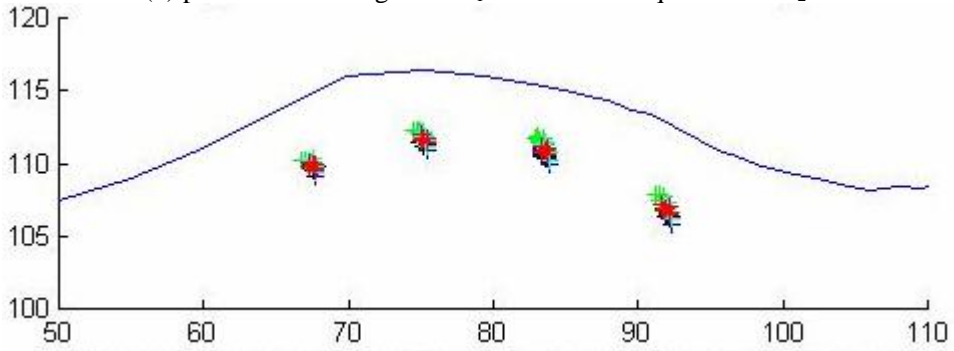
Figure 7.24 Simulations pour l'hypothèse de planification de voisinage avec exécution selon Öhman.
(L'exécution selon Öhman avec un décalage de 15ms de l'initiation du geste consonantique)
(Voir Figure 7.21 pour les détails)



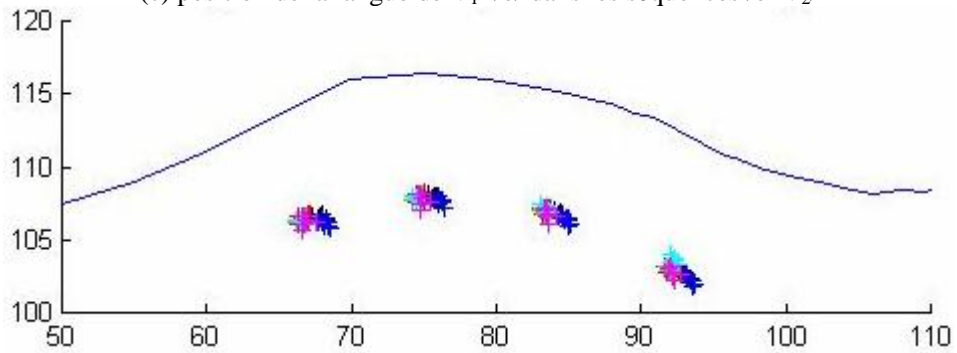
(a) position de la langue de /k/ dans les séquences $/V_1kV_2/$



(b) position de la langue de $V_1=/i/$ dans les séquences $/ikV_2/$



(c) position de la langue de $V_1=/e/$ dans les séquences $/ekV_2/$



(d) position de la langue de $V_1=/\epsilon/$ dans les séquences $/\epsilon kV_2/$

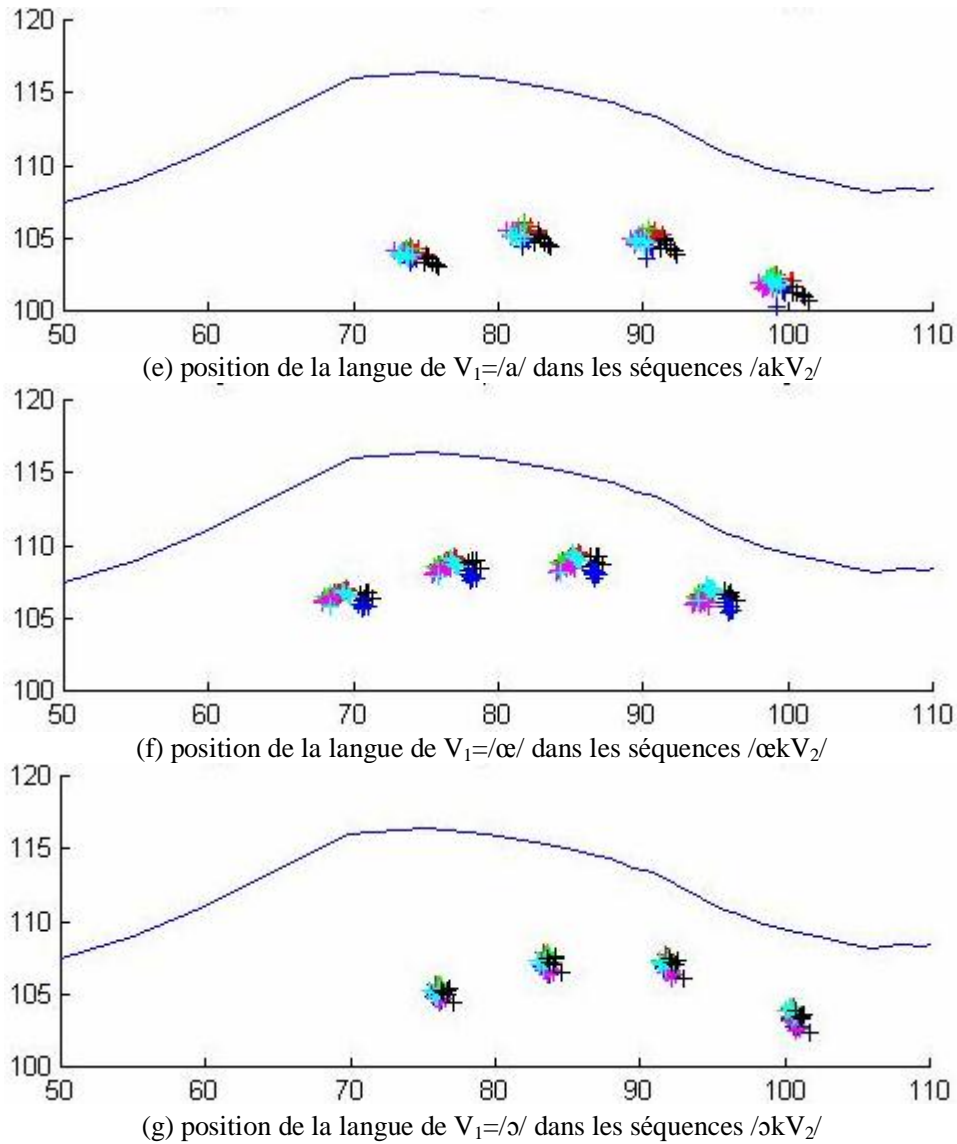


Figure 7.25 Simulations pour l'hypothèse de planification du voisinage avec l'exécution selon Öhman.
(L'exécution selon Öhman avec un décalage de 50ms de l'initiation du geste consonantique)
(Voir Figure 7.21 pour les détails)

Résumé :

Nos résultats des simulations montrent que, l'hypothèse de planification de type voisinage avec exécution séquentielle est conforme aux données articulatoires des sujets français. La voyelle V_2 influence à la fois la voyelle V_1 et la consonne C dans la séquence entière V_1CV_2 . L'hypothèse de planification syllabique avec exécution séquentielle est cohérente avec les données articulatoires des sujets chinois. Nous avons observé une influence de la voyelle V_2 sur la consonne C , mais pas sur la voyelle V_1 . C'est-à-dire que cette influence de V_2 est strictement limitée à la syllabe CV_2 pour la séquence de type V_1CV_2 .

Cependant, nous avons également observé l'influence de V_2 sur la séquence entière de V_1CV_2 pour l'hypothèse de planification inspirée du modèle d'Öhman avec exécution selon Öhman. Mais en comparaison avec l'hypothèse de planification de voisinage avec exécution séquentielle, nous constatons que la consonne /k/ est systématiquement moins influencée par V_2 pour l'hypothèse de planification selon Öhman avec exécution selon Öhman. De plus, la voyelle V_1 est plus affectée par V_2 pour cette hypothèse. Donc l'influence anticipatoire sur la voyelle V_1 est beaucoup plus grande que celle sur la consonne C . Cette tendance est en contradiction avec les observations des données.

Chapitre 8 : Discussion

Dans la première partie de notre travail, nous avons étudié la coarticulation dans des séquences de type V_1CV_2 en français et en chinois. Notre démarche a été guidée par l'hypothèse que le statut de la syllabe dans la représentation linguistique de ces deux langues est différent (chapitre 2). Pour le chinois, la syllabe est considérée comme l'unité la plus fondamentale de la structure phonologique, tandis que cette contrainte semble être moins forte pour le français. Des données articulatoires ont été tout d'abord recueillies sur trois sujets français et trois sujets chinois par un système électromagnétique (EMMA), en même temps que le signal acoustique de parole. Le corpus était composé de 15 phrases V_1CV_2 où les voyelles utilisées étaient /a i u/ et les consonnes étaient soit la vélaire /k/ soit l'alvéolaire /t/. Puis ensuite nous avons analysé les influences dues à la voyelle V2 sur les phonèmes précédents que nous avons interprétées comme le résultat de stratégies d'anticipation. Nous avons procédé à une analyse d'ANOVA (Mesures répétées) pour chaque sujet et avons enfin comparé les comportements d'anticipation des sujets dans les deux groupes.

Dans la seconde partie, nous avons élaboré un modèle de planification de séquences de parole s'appuyant sur un contrôle optimal et nous l'avons testé sur un modèle biomécanique de la langue. Nous avons fait 8800 simulations afin de constituer une base de données de formes articulatoires à partir de laquelle il a été possible d'élaborer une description générale des relations entre l'espace de commandes motrices et l'espace acoustico-perceptif. Le modèle interne qui correspond à une représentation de ces relations entre ces deux espaces a été réalisé grâce à un modèle fondé sur des fonctions radiales de base. Nous avons exploité ce modèle pour planifier les séquences motrices et produire des séquences de parole. La planification de la séquence consiste à inférer les commandes motrices à partir d'une spécification des objectifs cibles à atteindre dans l'espace acoustico-perceptif. Nous avons réalisé cette inversion en minimisant les efforts du locuteur dans l'espace des commandes motrices et en ajoutant des contraintes supplémentaires dans l'espace acoustico-perceptif. Différentes hypothèses concernant la structure phonologique de la séquence planifiée et les schémas temporels d'exécution du mouvement ont été testées en exploitant cette modélisation.

Nous faisons ci-dessous une synthèse de nos observations, d'abord sur les données

articulatoires, ensuite sur les simulations exploitant notre modélisation, et enfin sur la comparaison de ces deux types de résultats.

Les données articulatoires nous ont montré que,

- Pour une anticipation non adjacente de V_2 sur V_1 dans les séquences $/V_1tV_2/$:

L'influence de la voyelle V_2 sur V_1 est souvent significative pour les trois sujets français. Les observations nous permettent de constater que l'articulation de V_1 dans les séquences V_1CV_2 anticipe l'articulation de V_2 en français. En effet, dans la majorité des cas, $V_2=/i/$ influence spécifiquement les positions T_2 et T_3 de V_1 , qui sont situées dans la région de constriction de $/i/$. Par conséquent, des positions plus hautes et/ou plus antérieures de T_2 et de T_3 sont conformes à articulation suivante de $/i/$. De la même façon que pour $V_2=/i/$, dans certains cas pour $V_2=/a/$, la langue de V_1 montre des positions basses pour T_2 et T_3 . Nous pouvons penser que ceci est dû au fait que l'articulation de V_1 est influencée par une voyelle basse de $V_2=/a/$.

Pour les sujets chinois, l'influence associée au changement de V_2 sur la position de langue de V_1 est moins grande. Pour la plupart des cas, elles ne sont pas significatives. Le sujet WS montre une influence sur la position horizontale de T_4 . Mais la langue est plus postérieure pour $V_2=/i/$ que pour $V_2=/u/$ et/ou $V_2=/a/$. Cela est en contradiction avec une hypothèse d'anticipation. Il ne semble pas que ce soit le résultat direct d'une stratégie anticipatoire de l'articulation suivante. Nous supposons que c'est la conséquence de l'influence de la consonne $/t/$, quand elle est prononcée devant la voyelle $/i/$ par ce sujet. Nous observons que la langue du sujet JW pour la consonne $/t/$ est également localisée plus en arrière pour $V_2=/i/$ que pour $V_2=/a/$ et $V_2=/u/$.

- Pour une anticipation non adjacente de V_2 sur V_1 dans les séquences $/V_1kV_2/$:

Les sujets français présentent également un certain nombre d'effets d'anticipation de la voyelle V_2 , mais ces effets pour les séquences $/V_1kV_2/$ sont moins systématiques d'un sujet à l'autre et moins grands que pour les séquences $/V_1tV_2/$. Les observations générales confirment l'hypothèse selon laquelle l'anticipation de la voyelle V_2 existe pendant la production de la voyelle V_1 pour les sujets français. Nous avons rapporté que ce n'est pas le cas pour les sujets chinois. Les sujets chinois SK et WS ne présentent aucun effet significatif. Le seul cas significatif est observé pour le sujet JW. Le sujet JW montre une influence au changement de

V₂ sur T₂ et T₃. Les différences observées entre les séquences /V1kV2/ et /V1tV2/ ne semblent donc pas traduire de différences dans les stratégies de planification sous-jacentes. Elles sont vraisemblablement à mettre sur le compte des contraintes articulatoires plus fortes existant sur l'ensemble du corps de la langue pour la production de la consonne vélaire (cf plus loin).

- Pour une anticipation adjacente de V₂ sur C=/t/:

Pour tous les six sujets, nous avons observé une influence significative et très remarquable associée à V₂ sur la position verticale de T₂, T₃. La quantité d'influence liée aux changements de V₂ est aussi grande pour les sujets chinois que pour les sujets français. Toutes les différences significatives sont compatibles avec une anticipation de la voyelle V₂ sur la consonne C.

Nous avons constaté que le capteur T₁ est beaucoup moins sensible au changement de V₂ que les trois autres capteurs pour tous les six sujets dans nos données. Cela conforme le fait que l'apex de la langue est crucial pour la production de la consonne occlusive alvéolaire. La position de l'apex de la langue est prioritairement recrutée par la réquisition de la consonne. L'influence de la voyelle est moins forte sur cet articulateur.

En fait, pour les sujets chinois, dans la majorité des cas, ces données soutiennent l'hypothèse selon laquelle V₂ influence l'articulation de la consonne précédente C, mais d'une manière plus complexe. Par exemple, les trois sujets chinois montrent une position verticale de T₄ plus haute pour V₂=/i/ que pour V₂=/u/ et pour V₂=/a/ dans toutes les séquences. Cela est en contradiction avec une stratégie d'anticipation. Cette observation suggère que /t/ serait spécifié différemment selon la voyelle qui suit, et que ces spécifications correspondraient plus à la programmation globale de l'entité /tV₂/ qu'à une stratégie d'anticipation d'un phonème dans un autre. De plus, pour le sujet JW, nous avons observé que l'influence de la voyelle V₂ sur la voyelle V₁ ne peut pas être interprétée comme le résultat d'une anticipation de V₂ dans V₁ (voir une discussion pour la séquence /Vti/ dans le chapitre 6). À la lumière des positions articulatoires mesurées pour la consonne /t/ pour ce sujet, il semblerait qu'il s'agisse plutôt de l'influence de l'articulation cette consonne, ou encore d'une stratégie d'anticipation de la syllabe /tV₂/. En d'autres termes, selon les analyses faites sur les sujets chinois, il semble que leur planification de la séquence /tV₂/ exploiterait non seulement des stratégies anticipatoires, mais qu'elles prendraient en considération également la syllabe dans son ensemble pour

déterminer les cibles articulatoires spécifiques, et que cette entité globale serait à son tour intégrée dans la planification de la voyelle précédente.

En position horizontale, nous avons observé également des influences significatives associées à V_2 . Certaines influences sont en contradiction avec l'effet d'anticipation pour les six sujets. Cela pourrait être expliqué par les contraintes associées à la production de la consonne occlusive alvéolaire, qui exigent une occlusion dans la partie très antérieure du conduit vocal. Ces données montrent aussi bien les enjeux de la compétition entre le critère orienté vers le locuteur ("speaker oriented") et le critère orienté vers l'auditeur ("listener oriented") dans la production de la parole (Lindblom, 1990). Pour maintenir les caractéristiques articulatoires, cruciales dans un but perceptif, la consonne /t/ doit être moins anticipée en position horizontale. Donc les observations de tous nos sujets français et chinois reflètent une stratégie d'anticipation dans la direction verticale, et une exigence perceptive pour l'auditeur dans la direction horizontale pour la consonne /t/.

- Pour une anticipation adjacente de V_2 sur $C=/k/$:

Nous avons observé des variabilités significatives associées à V_2 en position horizontale pour les six sujets. Toutes ces influences significatives sont compatibles avec une stratégie d'anticipation. La position de la langue est plus en arrière pour $V_2=/u/$ que pour $V_2=/a/$. Des variabilités significatives en position verticale sont aussi observées, mais elles sont en contradiction avec une anticipation pour la plupart des cas. Ceci pourrait être expliqué par la combinaison de la stratégie anticipatoire du locuteur orienté, les contraintes perceptives orientées auditeur et les contraintes physiques : la perception demande une production de l'occlusion dans la partie région vélo-palatale pour /k/ et l'anticipation doit générer une position plus haute et plus postérieure pour la voyelle suivante /u/. Les positions plus postérieures sont observées, mais la forme du palais dans la région vélo-palatale limite l'élévation de la langue et impose la position verticale actuelle de la langue. Nous retrouvons que les influences observées dans la direction horizontale reflètent la stratégie d'anticipation et celles dans la direction verticale reflètent la morphologie du conduit vocal spécifique au locuteur.

- Pour les contraintes articulatoires :

Nous avons observé que, pour les six sujets, la quantité de variabilité de V_1 associée au changement de V_2 est fonction de la voyelle V_1 elle-même et de la consonne C. La voyelle

$V_1=i/$ semble être en général moins sensible à la variation de V_2 . Ceci confirme l'hypothèse selon laquelle $/i/$ est plus contrainte dans le domaine articulaire que les autres voyelles (Recasens 1984, 1987). Nous avons également observé que l'anticipation de V_2 sur V_1 est plus grande dans les séquences $/V_1tV_2/$ que dans les séquences $/V_1kV_2/$. Ceci confirme l'hypothèse du modèle DAC de Recasens qui suggère que la coarticulation de voyelle à voyelle dans la séquence VCV est inversement reliée au degré de contrainte articulaire de la consonne intervocalique sur l'activité du dos de la langue. En ce qui concerne l'anticipation adjacente de V_2 sur C, nos observations montrent que la position de langue de la consonne $/k/$ est beaucoup moins sensible au changement de V_2 que celle de la consonne $/t/$. Cette observation confirme que la consonne $/k/$ impose plus de contraintes pour le mouvement du corps de langue et que l'effet de voyelle V_2 sur la consonne C varie inversement avec le degré de contrainte s'appliquant sur la consonne.

Des différences sont observées pour l'anticipation des séquences V_1CV_2 en français et en chinois. Pour les sujets français, la voyelle V_2 influence la voyelle V_1 et la consonne C. Pour les sujets chinois, nous avons observé l'influence de voyelle V_2 sur la consonne C, et non pas sur la voyelle V_1 .

Une première explication possible pour les différences d'anticipation de V_2 à V_1 pourrait être liée aux systèmes vocaliques des deux langues. Il est suggéré que la densité du système vocalique pourrait influencer la variabilité associée à la coarticulation de Voyelle à Voyelle (Manuel, 1990) et la variabilité d'une répétition à l'autre des voyelles (Mooshammer *et al.*, 2004) : plus le système vocalique est dense, plus la variabilité est petite.

Nous avons proposé, dans le chapitre 2, une description simplifiée du système vocalique du chinois et du français. Selon une description phonologique classique, il existe cinq voyelles, $/i, y, u, a, \text{ə}/$ en chinois (Deng et Dang, 2007 ; Duanmu, 2002). Mais, la situation est classiquement décrite comme étant plus complexe à cause d'un certain nombre d'allophones émergeant dans différents contextes phonétiques. En français, le système des voyelles est composé de 11 voyelles orales monophthongues (Calliope, 1989). Si nous comparons le système vocalique du chinois et du français, malgré le fait que la description du système vocalique du chinois est toujours controversée, il est possible de dire que le chinois a tout au plus autant de voyelles que le français. De ce point de vue, si la densité du système vocalique influence la variabilité due à la coarticulation, la variabilité de V_1 en chinois devrait être plus

grande ou égale à celle du français. Nous observons le contraire dans nos données. Un certain nombre d'études ont cependant minimisé le rôle potentiel de la densité du système des voyelles sur la variabilité. Bradlow (1995), par exemple, a découvert que les voyelles dans un petit système (par exemple l'Espagnol avec un système de 5 voyelles) ne varient pas plus que les voyelles dans un grand système comme l'anglais et le français (les systèmes avec une dizaine de voyelles) (voir également Meunier, 2003). En outre, une étude sur la coarticulation de voyelle à voyelle en cantonais (système de 8 voyelles) et en mandarin chinois (considéré ici comme un système de 5 voyelles) a suggéré que la densité du système vocalique n'influence pas la coarticulation dans ces deux langues (Mok et Hawkins, 2004). Donc, nous pouvons rejeter la possibilité de cette explication de l'influence du système vocalique sur la coarticulation de voyelle à voyelle.

Une autre explication possible pourrait être liée au temps de transition pour atteindre la cible du phonème. Il a été suggéré que l'anticipation diminue quand la durée de transition entre les cibles augmente (Abry et Lallouache 1996). Donc les durées de la transition V_1-V_2 et de $C-V_2$ peuvent influencer la coarticulation de voyelle à voyelle et de voyelle à consonne.

Nous constatons que dans nos données, les consonnes sont produites de manière différente dans le groupe chinois et le groupe français. En effet, au lieu de prononcer les consonnes occlusives /t/ et /k/, les sujets chinois ont produit plutôt des consonnes occlusives aspirées /t^h/ et /k^h/. Les consonnes occlusives aspirées sont plus longues que les consonnes occlusives non aspirées, et ceci pourrait induire un plus grand intervalle de V_1 à V_2 et de C à V_2 qui pourrait donc induire moins de coarticulation. Pour exclure cette possibilité, nous avons testé les distributions de la durée V_1-V_2 et de la durée $C-V_2$ pour six sujets (chapitre 6).

L'analyse de la durée V_1-V_2 et $C-V_2$ nous a montré que :

Le sujet français CV qui a des durées très courtes se distingue significativement des autres sujets dans toutes les séquences. Les sujets chinois SK, WS et le sujet français AV qui ne se montrent pas significativement différentes dans la plupart des cas peuvent former un groupe. Ils ont des durées plus longues que le sujet CV. Le sujet chinois JW et le sujet français PB qui ont des durées plus longues forment un groupe, et significativement se distinguent des autres sujets. Donc, il n'y a pas de séparation significative dans les mesures de durée pour les deux groupes de langue. De plus, si nous comparons la durée de transition V_1-V_2 avec l'observation de l'anticipation V_1-V_2 , nous avons trouvé que : parmi les sujets français, CV qui a des durées plus courtes ne semble pas avoir plus d'anticipation que le sujet PB qui a des durées plus

longues. Donc dans nos données, il n'y a donc pas de relation entre l'anticipation et la durée à atteindre la cible. La différence observée dans l'influence de V_2 sur V_1 entre les deux groupes de sujets ne peut donc pas être expliquée par la durée de transition entre les cibles des phonèmes.

Nous interprétons plutôt nos observations par les différentes stratégies anticipatoires de coarticulation des sujets français et des sujets chinois comme les conséquences du statut respectif de la syllabe dans leurs langues.

Il a été suggéré (Levelt, 1994) que, dans toutes les langues, la parole serait basée sur l'existence d'un syllabaire mental ("*mental syllabary*"), dans laquelle les locuteurs stockeraient des syllabes phonologiques à un niveau particulier de représentation. Ce syllabaire servirait de mécanisme de base pour traduire les représentations phonologiques abstraites dans les descriptions phonétiques dépendantes du contexte, ce qui serait le point de départ de la génération des gestes de la parole et des signaux acoustiques.

Comme nous l'avons dit dans le deuxième chapitre, pour le chinois, la syllabe est classiquement considérée comme l'unité structurante phonologique la plus importante. C'est une caractéristique très forte de cette langue. Le français est également classiquement considéré comme une langue à base syllabique. Mais, la force de la syllabe au niveau phonologique en français semble être moins forte que celle en chinois (voir dans le chapitre 2).

Nos résultats ont montré que pour les sujets français, la voyelle V_2 influence la séquence entière de V_1CV_2 , tandis qu'elle est limitée dans la syllabe CV_2 pour les sujets chinois. Cela signifie que la planification de la parole en français peut s'étendre au-delà de la taille de la syllabe alors que pour le chinois la planification est limitée dans la syllabe CV_2 .

De plus, pour les sujets chinois, nous avons observé deux cas qui sont en contradiction avec une stratégie d'anticipation, et qui sont difficiles à interpréter. 1) la position verticale du T_4 de la langue de consonne /t/ (elle est plus haute pour $V_2=/i/$ que pour $V_2=/a/$ et pour $V_2=/u/$ dans toutes les séquences /VtV/). 2) le sujet chinois JW montre que sur la direction antérieure postérieure, la position de la langue de V_1 est plus en arrière quand la voyelle V_2 est une voyelle antérieure /i/. Donc il semble qu'en chinois la planification des séquences CV_2 ne consiste pas simplement en une optimisation de l'articulation entre deux phonèmes. D'autres critères spécifiques à l'unité CV_2 , semblent également jouer un rôle. La production d'une syllabe CV est spécifiquement organisée et commandée également comme une unité entière avec ses règles spécifiques incluses dans l'anticipation.

Nos observations pour lesquelles l'anticipation des sujets chinois est limitée à la syllabe sont cohérent avec les travaux dans le domaine de la coarticulation tonale. Phonologiquement, pour les langues à tons, la propagation (spreading) des tons de gauche à droite est assez commune tandis que la propagation des tons de droite à gauche est très rare (Akinlabi et Liberman, 2001). Cette asymétrie dans la propagation phonologique des tons correspond à une asymétrie dans la direction de la coarticulation tonale. La coarticulation rétentrice est généralement beaucoup plus forte que la coarticulation anticipatoire. Ainsi, beaucoup de langues à ton montrent une coarticulation anticipatoire minimale, par exemple, le cantonais (Li *et al.*, 2004), le mandarin (Xu, 1997), le thaï (Gandour *et al.*, 1994) et le vietnamiens (Brunelle, 2003). Pour le mandarin, Xu (1997) a observé que la coarticulation rétentrice et la coarticulation anticipatoire se différencient en amplitude et en nature pour les séquences dissyllabiques CV₁CV₂. Pour la coarticulation rétentrice, la valeur initiale de F₀ d'un ton est assimilée à la valeur de la fin du ton précédent. Cette influence peut être observée au moins aux deux tiers de la voyelle de la syllabe suivante et parfois même à la fin de la syllabe suivante. En revanche, pour l'anticipation, un ton bas en début de syllabe augmente la F₀ du ton précédent plutôt que ne l'abaisse. Cette asymétrie forte de la coarticulation tonale vers la coarticulation rétentrice peut être liée aux contraintes articulatoires inertielles associées à d'autres contraintes comme la position du ton dans la syllabe. (Flemming, 2007). Selon Flemming, le mouvement mélodique serait en effet réalisé sur les voyelles et sur toutes les consonnes sonores du coda parce que ces sons ont l'intensité la plus forte. Dans cette perspective, il serait plus efficace de réaliser les cibles d'un ton sur la rime d'une syllabe, plus sonore, que sur l'onset d'une syllabe. Du fait des contraintes inertielles du système phonatoire, la transition vers le ton suivant devrait obligatoirement se faire sur l'onset de la syllabe suivante, ce qui expliquerait la propagation du ton précédent au-delà des limites de la syllabes. (Flemming, 2007). Les contraintes physiques expliqueraient donc que la coarticulation rétentrice dépasse les limites de la syllabe. Par contre, tout comme le suggèrent nos résultats, aucune trace de coarticulation anticipatoire n'est noté au-delà des limites de la syllabe.

Nos observations pour les sujets chinois sont compatibles avec les propositions de Kozhevnikov et Chistovich (1965) au sujet du rôle principal de la syllabe dans l'organisation de la parole. Pour les sujets français nos observations confirment des modèles de coarticulation comme le modèle d'Öhman, le modèle de MEM (Abry et Lallouache, 1995) ou les modèles optimaux de la planification proposés par Jordan (1990), Kawato *et al.* (1990),

Perkell *et al.* (2000) ou Perrier *et al.* (2005), dans lesquels tous prennent en compte des séquences de phonèmes plus longues que la syllabe CV.

Les simulations d'une modélisation du contrôle moteur:

Différentes hypothèses ont été testées sur la planification et sur l'exécution du mouvement en utilisant notre modélisation de contrôle. Les résultats des simulations ont été analysés statistiquement par l'ANOVA (Mesures Répétées).

Les simulations nous ont montré que

- Pour une hypothèse de planification de voisinage associée à une exécution séquentielle :

L'hypothèse de planification de voisinage est une planification globale de rapprochement de voisinage. Elle prend en compte tous les phonèmes de manière équivalente. La planification est donc définie par le rapprochement global de toutes les cibles de la séquence V_1 -C- V_2 . Les simulations avec cette hypothèse de voisinage associée à une exécution séquentielle montrent des effets d'anticipation de la voyelle V_2 sur la position de la position atteinte de première voyelle V_1 et celle de la consonne /k/.

- Pour une hypothèse de planification syllabique associée à une exécution séquentielle :

L'hypothèse syllabique est une planification qui donne un statut spécifique à la syllabe CV₂. Les simulations avec cette hypothèse syllabique associée à une exécution séquentielle montrent que la position atteinte pour la voyelle /a/ n'est pas affectée par la voyelle V_2 ; Par contre on observe une grande influence de V_2 pour la consonne /k/.

- Pour une hypothèse de planification inspirée du modèle d'Öhman associée à une exécution selon Öhman :

L'hypothèse inspirée du modèle d'Öhman donne un statut spécifique à la transition vocalique V_1V_2 pour la planification. Les simulations avec cette hypothèse associée à une exécution à la Öhman montrent de plus larges effets d'anticipation de la voyelle V_2 . La position de la voyelle /a/ est plus affectée par la voyelle V_2 . Notons toutefois que cette hypothèse montre que l'influence de V_2 sur la position de la voyelle V_1 est plus grande que celle sur la position de la consonne /k/ par rapport à l'hypothèse de voisinage associée à une exécution séquentielle.

- Pour une hypothèse de planification de voisinage associée à une exécution selon Öhman,

Les simulations montrent que pour la même hypothèse de planification, dans les conditions normales de parole, le modèle d'exécution selon Öhman ne se différencie pas d'un

modèle d'exécution séquentielle sur la position atteinte par la langue de V_1 et C. Nous observons seulement un décalage temporel du mouvement vers la consonne entre les deux modèles d'exécution.

Une comparaison entre les données articulatoires et les simulations :

L'hypothèse de planification de voisinage avec exécution séquentielle est conforme aux données des sujets français : La position de la voyelle V_1 est affectée par la voyelle V_2 . La consonne C est ainsi influencée par le changement de la voyelle V_2 . Cependant, l'hypothèse de planification selon Öhman avec exécution selon Öhman montre également un large effet d'anticipation de V_2 à V_1 et V_2 à C, mais cette hypothèse montre que l'influence de V_2 sur la position de V_1 est plus grande que celle sur la position de la consonne C au changement de V_2 . C'est en contradiction avec l'observation des données.

L'hypothèse de planification syllabique avec exécution séquentielle est cohérente avec les données articulatoires des sujets chinois : nous n'avons pas observé d'influence de la voyelle V_2 sur la position de V_1 , mais nous avons observé une anticipation beaucoup plus claire de la voyelle V_2 sur la consonne C.

Pour l'exécution séquentielle, nous avons également testé les simulations sans planification. Elles ne sont pas réalistes car elles ne montrent aucun effet d'anticipation de la position d'un phonème sur le précédent. Ceci contredit les observations faites pour tous les locuteurs. L'hypothèse sans planification nous a montré que le modèle purement physique ne génère pas le phénomène d'anticipation.

La comparaison entre les simulations et les données articulatoires est conforme à notre hypothèse selon laquelle, en français, la planification prend en compte la séquence entière de V_1CV_2 tandis que, en chinois, la planification est limitée dans la syllabe CV_2 .

Notre étude des données articulatoires a mis en place un corpus constitué de répétitions de séquences $/V_1CV_2/$; nous pensons que les effets d'anticipation seraient plus grands en contexte de parole continue. Nous pouvons nous attendre à observer un phénomène d'anticipation plus important en parole spontanée. De plus, le corpus a été élaboré afin d'avoir un environnement conforme de coarticulation dans les deux langues. Toutes les séquences $/V_1CV_2/$ que nous avons prises n'ont aucun sens ni en français ni en chinois. Nous pensons également que les faits pourraient être légèrement différents dans une contexte de phrase

ayant un sens.

Nous avons également travaillé dans cette étude sur deux consonnes : une consonne alvéolaire /t/ et une vélaire /k/. Ce sont les deux consonnes qui ont le plus de contraintes sur l'apex de la langue et le dos de la langue. L'effet d'anticipation de V_2 à V_1 et de V_2 à C est lié aux contraintes de la consonne, comme Recasens l'a proposé dans le modèle DAC. Nous pensons que cet effet serait plus varié avec une plus grande variété de consonnes, et avec plus grande combinaison de voyelles et de consonnes.

Pour la modélisation, nous avons défini les cibles dans un espace acoustico-perceptif en faisant l'hypothèse que tous les sons, y compris les consonnes, sont spécifiés par des caractéristiques particulières dans l'espace (F_1 , F_2) et que chaque phonème a une région cible définie par des centres F_{c1} , F_{c2} et des écart-types σ_{F1} , σ_{F2} . Pour définir ces paramètres pour les voyelles, nous nous sommes largement inspirés de données publiées dans l'ouvrage de Calliope sur le français. Mais les variabilités des données de Calliope sont calculées pour une dizaine de sujets français. Ils ne correspondent pas à une variabilité intra-sujet. Nous avons envie de définir la cible perceptive sur un seul sujet français. Nous pensions que les cibles seraient réduites si nous faisons une définition de la variabilité sur un seul sujet.

Les simulations dans notre étude ont été générées par un échantillonnage de l'espace des variables de commande sur une distribution uniforme des paramètres de commande autour de la forme neutre de la langue. Mais la distribution de simulation dans le plan formantique (F_1 , F_2) n'est pas uniforme à cause de la relation non linéaire entre l'espace articulatoire et l'espace acoustique. Notre espace perceptif-acoustique n'est pas bien couvert par les voyelles dans la partie centrale et pour la consonne /k/ dans la partie haute. Une représentation beaucoup plus dense se trouve sur l'axe /i, e, ε, a/. Cette distribution non uniforme nous pose le problème de l'apprentissage du modèle interne. Nous pensons que le modèle interne aura une relation plus stable avec beaucoup plus de simulations qui couvrent bien l'espace acoustico-perceptif.

Conclusion

Nous avons effectué une étude expérimentale et une modélisation du contrôle de la planification de la parole dans le but d'étudier les stratégies de coarticulation anticipatoires et leur lien potentiel avec la structure phonologique de la langue. Le premier travail a permis de collecter des données articulatoires grâce au système EMMA. Nous avons travaillé sur la coarticulation anticipatoires pour deux langues, le chinois et le français, pour qui il est proposé que le statut de syllabe soit différent au niveau phonologique. Nous avons proposé une méthode d'étiquetage acoustique et d'étiquetage articulatoire. Pour la séquence V_1CV_2 , nous avons effectué une étude de l'influence de V_2 sur V_1 et de l'influence de V_2 sur C pour chaque sujet et avons comparé les comportements des sujets dans deux groupes de langue. Différentes stratégies ont été observées. Nos résultats ont montré que pour les sujets français, la voyelle V_2 influence la séquence entière de V_1CV_2 , tandis qu'elle est limitée à la syllabe CV_2 pour les sujets chinois. Cela signifie que la planification de la parole en français peut s'étendre au-delà de la taille de syllabe alors que ce n'est pas le cas pour le chinois. Ces résultats sont interprétés comme les conséquences du statut respectif de la syllabe dans leurs langues. Ils ont démontré l'aspect des contraintes linguistiques de la langue sur des stratégies articulatoires de planification dans la parole.

Au-delà de ces observations phénoménologiques, il est intéressant de savoir plus concrètement comment les mécanismes de planification fonctionnent et comment nous pouvons contrôler les différentes stratégies dans le système moteur. C'est là l'intérêt du modèle. Il permet de tester quantitativement des hypothèses de planification sur le contrôle de la parole en comparaison avec les données.

Pour cela, nous avons établi une démarche de contrôle optimal de la planification pour un modèle interne. Le modèle interne a décrit des relations statiques entre les commandes motrices et les caractéristiques acoustiques de nos données de simulations. Il correspond à une représentation du système nerveux central. Nous faisons une inversion pour trouver les commandes motrices associées à la génération des caractéristiques spectrales cibles souhaitées grâce au modèle interne. Cette inversion est réalisée par des solutions optimales qui intègrent à la fois des contraintes orientées vers le locuteur (minimisation d'un « effort »)

et des contraintes orientées vers l'auditeur (maximisation de l'efficacité perceptive du son à la cible). Nous avons testé dans cette étape des solutions optimales sur des hypothèses de planification des commandes cibles : une hypothèse globale de planification de voisinage, une hypothèse de planification syllabique et une hypothèse inspirée du modèle d'Öhman. En fin de planification, les cibles planifiées d'une séquence sont prêtes. Mais la production de séquences est également influencée par les caractéristiques temporelles du système de génération du mouvement. Pour savoir comment générer les mouvements entre cibles, nous avons testé deux types d'exécution du mouvement, l'exécution séquentielle et l'exécution selon Öhman. Nous avons comparé les résultats des simulations de séquences VCV et ceux des données articulatoires. Les résultats ont suggéré que les stratégies anticipatoires du locuteur chinois pourraient être les résultats d'une planification syllabique dans laquelle la planification donne un statut spécifique à la syllabe, tandis que le locuteur français pourrait faire une stratégie globale du voisinage dans laquelle la planification concerne la séquence entière de VCV.

Nous avons cependant indiqué que le nombre de phonèmes dans les simulations et le nombre de données n'étaient pas suffisants pour une évaluation complète. Nos conclusions doivent bien sûr être approfondies en augmentant le nombre de simulations par la prise en compte d'une plus grande variété de consonnes, et d'une plus grande combinaison de voyelles et de consonnes. De nombreux domaines de recherche nous intéressant restent à explorer. Nous souhaitons étudier les données articulatoires au moyen d'EMMA en 3D, ce qui nous permettrait d'observer la langue dans sa globalité. Pour découvrir l'exécution du mouvement de la langue, nous souhaiterions étudier les données cinématiques d'EMMA. Nous aimerions également comparer les instants de début et de fin des mouvements vocaliques V_1 - V_2 , les instants de début et de fin des mouvements consonantiques V_1 -C et C- V_2 , ainsi que, pour chacun de ces mouvements, la position du maximum de vitesse. Ces données cinématiques pourraient nous donner la possibilité d'étudier la stratégie d'exécution du mouvement du locuteur et de la comparer avec les simulations du modèle.

Bibliographie:

- Abry, C. & Lallouache, T.-M. (1991) Does increasing representational complexity lead to more speech variability? On explaining some French data by current models of anticipatory rounding. In O. Engstrand. & C. Kylander. (eds), *Current Phonetic Research Paradigms: Implications for Speech Motor Control*, Perilus, XIV, University of Stockholm, Sweden, pp. 1-5.
- Abry, C. & Lallouache, T.-M. (1995) Modelling lip constriction anticipatory behavior for rounding in French with the MEM (Movement Expansion Model). *Proceedings of the 13th International Congress of Phonetic Sciences*, 4, Stockholm, Sweden, pp. 152-155.
- Abry, C. & Lallouache, T.-M. (1996) Le MEM: Un modèle d'anticipation paramétrable par le locuteur. Données sur l'arrondissement du français. *Bulletin de la Communication Parlée*, 3, Institut de la Communication Parlée, Grenoble, France, pp. 85-89.
- Abry, C. & Perrier, P. (1997) Le contrôle des mouvements audibles et visibles dans la parole. In *Fondements et perspectives en traitement automatique de la parole*, Paris : Hachette, pp. 55-80.
- Akinlabi, A. & Liberman, B. (2001) Tonal complexes and tonal alignment. *Proceedings of the North East Linguistic Society*, 31, Minjoo Kim and Uri Strauss (eds), GLSA, MA, pp. 1-20.
- Badin, P. & Fant, G. (1984) Notes on vocal tract computations. *STL QPSR*, 2-3, KHT, Stockholm, pp. 53-108.
- Bailly, G., Jordan, M., Mantakas, M., Schwartz, J.-L., Bach, M. & Olesen, M. (1990) Simulation of vocalic gestures using an articulatory model driven by a sequential neural network. *Journal of the Acoustical Society of America*, 87, S105.
- Bailly, G., Laboissière, R. & Schwartz, J.-L. (1991) Formant trajectories as audible gestures: An alternative for speech synthesis. *Journal of Phonetics*, 19, pp. 9-23.
- Bao, H.Q. & Zheng, Y.L. (2001) A tentative statistic analysis of the EPG data of speech in standard Chinese. *Modern Phonetics of the New Century: Proceedings of the 5th National Conference on Modern Phonetics*, Beijing: Qinghua Press, pp. 9-17.
- Bell-Berti, F. & Harris, K.S. (1979) Anticipatory Coarticulation: Some implications from a study of lip rounding. *Journal of the Acoustical Society of America*, 65, pp. 1268-1270.
- Bell-Berti, F. & Harris, K.S. (1981) A temporal model of speech production. *Phonetica*, 38, pp. 9-20.
- Bell-Berti, F. & Krakow, R.A. (1991) Anticipatory velar lowering: A coproduction account. *Journal of the Acoustical Society of America*, 90 (1), pp. 112-123.

- Benguerel, A.-P. (1974) Nasal Airflow Patterns and Velar Coarticulation in French. *Speech Communication Seminar Proceedings, Vol. II*, Stockholm, Sweden: Almqvist and Wiksell, pp. 105-112.
- Benguerel, A.-P. & Cowan, H.A. (1974) Coarticulation of upper lip protrusion in French. *Phonetica*, 30, pp. 41-55.
- Benguerel, A.-P., Hirose, H., Sawashima, M. & Ushijima, T. (1977a) Velar coarticulation in French: a fiberoptic study. *Journal of Phonetics* 5, pp. 149-158.
- Benguerel, A.-P., Hirose, H., Sawashima, M. & Ushijima, T. (1977b) Velar coarticulation in French: an electromyographic study. *Journal of Phonetics* 5, pp. 159-167.
- Bensaber, K. (1998). *Conception et réalisation d'une base de données cinématiques : Application à l'étude de la coarticulation*. Thèse de doctorat en Sciences du Langage, Université Stendhal, Grenoble, 130 p.
- Boë, L.-J. & Perrier, P. (1990) Comment on "distinctive regions and modes: A new theory of speech production" by M. Mrayati, R. Carré and B. Guerin. *Speech Communication*, 9, pp. 217-230.
- Boyce, S.E., Krakow, R.A., Bell-Berti, F. & Gelfer, C.E. (1990) Converging sources of evidence for dissecting articulatory movements into core gestures. *Journal of Phonetics*, 18, pp. 173-188.
- Boyce, S.E., Krakow, R.A. & Bell-Berti, F. (1991) Phonological underspecification and speech motor organisation. *Haskins Laboratories Status Report on Speech Research, SR-105/106*, Haskins Lab, New Haven, CT, USA, pp. 141-152.
- Bradlow, A.R. (1995) A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America*, 97, pp. 1916-1924.
- Brichler-Labaeye, C. (1970) *Les voyelles françaises : Mouvement et position articulatoires à la lumière de la cinéradiographie*. Paris : Klincksieck.
- Browman, C.P. & Goldstein, L. (1986) Towards an articulatory phonology. *Phonology Yearwork*, 3, pp. 219-251.
- Browman, C.P. & Goldstein, L. (1989) Articulatory gestures as phonological units. *Phonology*, 6, pp. 201-251.
- Browman, C.P. & Goldstein, L. (1992) Articulatory phonology: an overview. *Phonetica*, 49, pp. 155-180.
- Calliope (1989) *La parole et son traitement automatique*. Paris : Masson, 736 p.
- Carney, J.-P. & Moll, K. (1971) A cinefluorographic investigation of fricative consonant-vowel coarticulation. *Phonetica*, 23, pp. 193-202.

- Carré, R. (1991) Vowel-Vowel trajectories and region modelling. *Journal of Phonetics*, 19, pp. 433-443.
- Carré, R., Bourdeau, M. & Tubach, J.-P. (1995) Vowel-Vowel Production: The Distinctive Region Model (DRM) and Vowel Harmony. *Phonetica*, 52, pp. 205-214.
- Catford, J.C. (1977) *Fundamental Problems in Phonetics*. Edinburgh University Press, 278 p.
- Cathiard, M.-A. (1994) *La perception visuelle de l'anticipation des gestes vocaliques*. Thèse en psychologie cognitive. Université Pierre Mendès-France (Grenoble II), Grenoble.
- Chao, Y.R. (1928) *Studies of Wu Dialects*. Tsing hua University Press, China.
- Chao, Y.R. (1968) *A Grammar of Spoken Chinese*. Berkeley, CA: University of California Press.
- Chomsky, N. & Halle, M. (1968) *The Sound Pattern of English*. New York: Harper & Row.
- Content, A., Meunier, C., Kearns, R.K. & Frauenfelder, U.H. (2001a) Sequence detection in pseudowords in French: Where is the syllable effect? *Language and Cognitive Processes*, 16, pp. 609-636.
- Content, A., Kearns, R. K., & Frauenfelder, U. H. (2001b) Boundaries versus onsets in syllabic segmentation. *Journal of Memory and Language*, 45, pp. 177-199.
- Da, J. (2000) *Chinese text computing*. <http://lingua.mtsu.edu/chinese-computing/>. Murfreesboro, TN: Department of Foreign Languages and Literatures, Middle Tennessee State University.
- Dang, J., Honda, M. & Honda, K. (2004) Investigation of coarticulation in continuous speech in Japanese. *Acoustical Science and Technology*, 25(5), pp. 318-329.
- Dang, J. & Honda, K. (2004) Construction and control of a physiological articulatory model. *Journal of the Acoustical Society of America*, 115(2), pp. 853-870.
- Dang, J., Wei, J., Suzuki, T. & Perrier, P. (2005) Investigation and Modeling of Coarticulation during Speech. *Proceedings of Interspeech 2005*, Lisbon, Portugal, pp. 1025-1028.
- Dang, J., Akagi, M. & Honda, K. (2006) Communication between speech production and speech perception within the brain. Observation and simulation. *Journal of Computer Science and Technology*, 21(1), pp. 95-105.
- Delais-Roussarie, E. (1995) *Pour une approche parallèle de la structure prosodique*. Thèse de Doctorat en Sciences du Langage, Université Toulouse-le Mirail, 297 p.
- Delattre, P. (1939) Accent de mot et accent de groupe. *The French Review*, 13(2), pp. 141-146.
- Delattre, P., Liberman, A.M. & Cooper, F.S. (1955) Acoustic loci and transition cues for consonants. *Journal of the Acoustical Society of America*, 27, pp. 769-773.

- Delattre, P. (1940) Le mot est-il une entité phonétique en français? *Le Français Moderne*, 8(1), pp. 47-56.
- Delattre, P. (1953) Les modes phonétiques du français. *The French Review*, XXVII (I), pp. 59-63.
- Delattre, P. (1969) Syllabic features and phonic impression in English, German, French and Spanish. *Lingua* 22, pp. 160-175.
- Deng, L. & Dang, J. (2007) Speech Analysis: The Production-Perception Perspective. In Hai-Zhou Li & Chin-Hui Lee (eds), *Advances in Chinese Spoken Language Processing*, Publisher: World Scientific, New Jersey, pp. 3-32.
- Di Cristo, A. & Hirst, D. (1993) Rythme syllabique, rythme mélodique et représentation hiérarchique de la prosodie du français. *Travaux de l'Institut de Phonétique d'Aix*, 15, pp. 9-24.
- Duanmu, S. (2002) *The Phonology of Standard Chinese*. Oxford: Oxford University Press.
- Duanmu, S. (2006) Chinese (Mandarin): Phonology. *Encyclopedia of Language and Linguistics*, 2nd ed., Keith Brown (ed.), Oxford, UK: Elsevier Publishing House, pp. 351-355.
- Fagyal, Z., Nguyen, N. & Boula de Mareüil, P. (2003) From dilation to coarticulation: is there vowel harmony in French? *Studies in Linguistic Sciences*, 32, pp. 1-21.
- Fant, G. (1960) *Acoustic Theory of Speech Production*, The Hague, NL: Mouton.
- Farnetani, E., Hardcastle, W. & Marchal, A. (1989) Cross-language investigation of lingual coarticulatory processes using EPG. In *Proceedings of the Eurospeech '89*, Paris, pp. 429-432.
- Farnetani, E. (1990) VCV lingual coarticulation and its spatiotemporal domain. In W.J. Hardcastle & A. Marchal (eds), *Speech Production and Speech Modelling*, Kluwer Academic Publishes, pp. 93-130.
- Feldman, A.G. (1986) Once more on Equilibrium Point Hypothesis for Motor Control. *Journal of Motor Behaviour*, 18 (1), pp. 17-54.
- Flemming, E. (2007) The grammar of coarticulation. *Workshop de l'Association Francophone de la Communication Parlée: La coarticulation : Indices, Direction et Représentation*, Montpellier, France, pp. 23-26.
- Flemming, E. (2001) Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology*, 18(1), pp. 7-44.
- Fonagy, I. (1979) L'accent en français : accent probabilitaire. In I. Fonagy & P. Léon (éds), *L'accent en français contemporain*, *Studia Phonetica* 15, Paris : Didier, pp. 123-233.
- Fouché, P. (1952) *Phonétique historique du français*. Paris : Klincksieck.

- Fougeron, C. & Jun, S.A. (1998) Rate Effects on French Intonation: Phonetic Realization and Prosodic Organisation. *Journal of Phonetics*, 26(1), pp. 45-70.
- Fowler, C.A. (1977) *Timing Control in speech production*. Bloomington, IN: Indiana University Linguistics Club.
- Fowler, C.A. (1980) Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, 8, pp. 113-133.
- Fowler, C.A. (1986) An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14, pp. 3-28.
- Frauenfelder, U. & Nguyen, N. (2000) La reconnaissance des mots parlés. In J. A. Rondal & X. Seron (éds), *Troubles du Langage : Bases Théoriques, Diagnostic et Rééducation*, Bruxelles : Mardaga, pp. 213-240.
- Galván-Rdz, A. (1997) *Etudes dans le cadre de l'inversion acoustico-articulatoire : Amélioration d'un modèle articulatoire, normalisation du locuteur et récupération du lieu de constriction des occlusives*. Thèse de l'Institut National Polytechnique de Grenoble.
- Goldstein, L. & Fowler, C.A. (2003) Articulatory phonology: A phonology for public language use. In Niels O. Schiller & Antjie S. Meyer (eds), *Phonetics and phonology in Language Comprehension and Production*, Berlin : Mouton de Gruyter, pp. 159-207.
- Gribble, P. L. & Ostry, D. J. (1999) Compensation for Interaction Torques During Single - and Multijoint Limb Movement. *Journal of Neurophysiology*, 82, pp. 2310-2326.
- Gribble, P.L., Ostry, D.J., Sanguineti, V. & Laboissière, R. (1998) Are complex control signals required for human arm movement? *Journal of Neurophysiology*, 79, pp. 1409-1424.
- Grammont, M. (1938) *Traité pratique de prononciation française*, 9^e éd., Paris : Delagrave.
- Guenther, F.H. (1995) Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, 102(3), pp. 594-621.
- Guenther, F.H., Hampson, M. & Johnson, D. (1998) A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, 105, pp. 611-633.
- Guenther, F.H., Ghosh, S.S. & Tourville, J.A. (2006) Neural modelling and imaging of the cortical interactions underling syllable production. *Brain and Language*, 96, pp. 280-301.
- Hamlet, S.L. & Stone, M.L. (1981) Pre-speech posturing of the mandible in relation to the jaw activity during speech. *Journal of Phonetics*, 9, pp. 425-436.
- Hardcastle, W. J. (1976) *Physiology of Speech Production*. An introduction for speech scientists. London: Academic Press.
- Hardcastle, W.J. (1985) Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication*, 4, pp. 247-263.

- Henke, W.L. (1966) *Dynamic articulatory model of speech production using computer simulation*. Ph.D. Thesis, MIT.
- Hirayama, M., Vatikiotis-Bateson, E., Kawato, M. & Jordan, M.I. (1992) Forward dynamics modeling of speech motor control using physiological data. In R.P. Lippman, J.E. Moody & D.S. Touretsky (eds), *Advances in Neural Information Processing Systems, 4*, San Mateo, CA: Morgan Kauffmann Publishers, pp. 191-198.
- Hoole, P., Nguyen, N. & Hardcastle, W.J. (1993) A comparative investigation of coarticulation in fricatives: electropalatographic, electromagnetic and acoustic data. *Language and Speech, 36*(2/3), pp. 235-260.
- Jordan, M.I. (1989) Serial order: A parallel distributed processing approach. In J. L. Elman and D. E. Rumelhart (eds), *Advances in Connectionist Theory: Speech*, Hillsdale, NJ: Lawrence Erlbaum Associates.
- Jordan, M.I. (1990). Motor Learning and the Degrees of Freedom Problem. In M. Jeannerod (ed.), *Attention and Performance*, Hillsdale, NJ: Erlbaum, pp. 796-836.
- Jordan, M.I. & Rumelhart, D.E. (1992) Forward Models: Supervised Learning with a Distal Teacher. *Cognitive Science, 16*, pp. 316-354.
- Jun, S.A. & Fougeron, C. (1995) The accentual phrase and the prosodic structure of French. *Proceedings 13th International Congress of Phonetic Sciences 2*, Kjell Elenius & Peter Branderud (eds), Stockholm, Sweden, pp. 722-725.
- Jun, S.A. & Fougeron, C. (2002) The Realizations of the Accentual Phrase in French Intonation. In J. Hualde (ed.), *Probus, 14, a special issue on Intonation in the Romance Languages*, pp. 147-172.
- Karlgren, B. (1949) *The Chinese Language: An essay on its Nature and History*. New York: The Ronald Press Company.
- Kawato, M., Maeda, Y., Uno, Y. & Suzuki, R. (1990) Trajectory formation of arm movement by cascade neural network model based on minimum torque-change criterion. *Biological Cybernetics, 62*, pp. 275-288.
- Kawato, M., Furukawa, K. & Suzuki, R. (1987) A hierarchical neural network model for control and learning of voluntary movement. *Biological Cybernetics, 57*, pp. 169-185.
- Keating, P. (1988a) Underspecification in phonetics. *Phonology, 5*, pp. 275-292.
- Keating, P. (1988b) The window model of coarticulation: articulatory evidence. *UCLA working papers in phonetics, 69*, pp. 3-29.
- Kelso, J.A.S. & Tuller, B. (1983) Compensatory articulation under conditions of reduced afferent information: a dynamic formulation. *Journal of Speech and Hearing Research, 26*, pp. 217-224.

- Kelso, J.A.S. & Tuller, B. (1984) Converging evidence in support of common dynamical principles for speech and movement coordinative. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology*, 246, R928-R935.
- Kelso, J. A. S., Saltzman, E. L. & Tuller, B. (1986) The dynamical theory on speech production: data and theory. *Journal of Phonetics*, 14, pp. 29-60.
- Kent, R.D. & Minifie, F.D. (1977) Coarticulation in recent speech production models. *Journal of the Acoustical Society of America*, 5, pp. 115-133.
- Khaligi, A. (1999) *Extraction des degrés de liberté d'un modèle biomécanique de la langue par analyses en composantes principales*. Mémoire de DEA en Signal-Image-Parole et Télécom. Institut National Polytechnique de Grenoble, France.
- Klein, C.A. & Huang, C. (1983) Review of pseudoinverse control for use with kinematically redundant manipulators. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-13, pp. 245-250.
- Kozhevnikov, V. & Chistovich, L. (1965). *Speech Articulation and Perception*, Joint Publication Research Service, Washington, D.C, pp. 104-118.
- Laboissière, R., Schwartz, J.-L. & Bailly, G. (1990) Motor control for speech skills: a connectionist approach. In Touretzky, D. S., Elman, J. L., Sejnowski, T. J. et Hinton, G. E. (eds), *Proceedings of Connectionist Models Summer School (CMSS)*, San Diego, USA, pp. 319-327.
- Laboissière, R. (1992) *Préliminaires pour une robotique de la communication parlée inversion et contrôle d'un modèle articulatoire du conduit vocal*. Thèse en Signal-Image-Parole et Télécom, Institut National Polytechnique de Grenoble, France.
- Levelt, W.J.M. & Wheeldon, L. (1994) Do speakers have access to a mental syllabary? *Cognition*, 50, pp. 239-269.
- Lieberman, A.M. & Mattingly, I.G. (1985) The motor theory of speech perception revisited. *Cognition*, 21, pp. 1-36.
- Lindblom, B. (1963) Spectrographic study of vowel production. *Journal of the Acoustical Society of America*, 35, pp. 1773-1781.
- Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In W.J. Hardcastle & A. Marchal (eds), *Speech Production and Speech Modelling*, The Netherlands: Kluwer Academic Publishers, pp. 403-439.
- Liu, F. (1924) *Experiments of Chinese Four Tones*. Qun-yi Book store.
- Löfqvist, A. & Yoshioka, H. (1981) Interarticulator programming in obstruent production. *Phonetica*, 38, pp. 21-34.

- Löfqvist, A. (1990) Speech as audible gestures. In W.J. Hardcastle & A. Marchal (eds), *Speech Production and Speech Modelling*, The Netherlands: Kluwer Academic Publishers, pp. 289-322.
- Lubker, J.F. & Lindgren, R. (1982) The perception effects of anticipatory coarticulation. In P. Hurme (ed.), *Speech Research*, Institut of Finnish language and communication, University of Jyväskylä, pp. 252-271.
- Lubker, J.F. & Gay, T. (1982) Anticipatory labial coarticulation: experimental, biological and linguistic variables. *Journal of the Acoustical Society of America*, vol 71(2), pp. 437-448.
- Ma, L. (2004) *Contribution à l'étude expérimentale et à la modélisation de la coarticulation en français et en mandarin dans les séquences Voyelle-Consonne-Voyelle*. Mémoire de DEA en Signal-Image-Parole et Télécom. Institut National Polytechnique de Grenoble, 36 p.
- Ma, L., Perrier, P & Dang, J. (2006) Anticipatory coarticulation: A crosslinguistic study of French and Mandarin speakers. *7th International Seminar on Speech Production*, Ubatuba, Brazil, pp. 151-158.
- Mackay, D. (1982) The problem of flexibility, fluency and speed-accuracy tradeoff. *Psychological Review*, 89, pp. 483-506.
- MacNeilage, P.F. (1972). Speech physiology. In J.H. Gilbert (ed.), *Speech and Cortical Functioning*, New York: Academic Press, pp. 1-72.
- MacNeilage, P. & DeClerk, J.L. (1969) On the motor control of coarticulation in CVC monosyllables. *Journal of the Acoustical Society of America*, 45, pp. 1217-1233.
- Maeda, S. (1979) Un modèle articulatoire de la langue avec des composantes linéaires. *Actes des 10èmes Journées d'Etude sur la Parole*, pp. 154-162.
- Maeda, S. (1990) Compensating articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model. In W.J. Hardcastle & A. Marchal (eds), *Speech Production and Speech Modelling*, The Netherlands: Kluwer Academic Publishers, pp. 131-150.
- Manuel, S & Krakow, R.A. (1984) Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratoires Status Report on Speech Research*, SR-77/78, Haskins Lab, New Haven, CT, USA, pp. 69-78.
- Manuel, S. (1990) The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America*, 88, pp. 1286-1298.
- Maraoui, M. (2004) *Etude préliminaire de la perception auditive des mouvements biologiques de la parole*. Mémoire de DEA en Sciences Cognitives, Institut National Polytechnique de Grenoble.

- Marchal, A. (1983) Coarticulatory pattern in stop sequences: EPG evidence. *Proceedings of the 10th International Congress of Phonetic Sciences*, Utrecht: Foris.
- Marchal, A. (1985) Description articulatoire et acoustique des groupes d'occlusives. *Travaux de l'Institut de Phonétique d'Aix-en-Provence 10*, pp. 13-61.
- Marchal, A. (1985) *L'électropalatographie : Contribution à l'étude de la coarticulation dans les groupes d'occlusives*. Thèse de doctorat d'état en Lettres et Sciences Humaines, Université de Nancy II, 984 p.
- Marchal, A. (1988) Coproduction: Evidence from EPG data. *Speech Communication*, 7, pp. 287-295.
- Marchal, A. & Hardcastle, W.J. (1993) ACCOR: Instrumentation and Database for the Cross-Language Study of Coarticulation. *Language and Speech*, 36, pp. 137-153.
- Marret, R. (2002) *Apprentissage des relations entre commandes musculaires et géométrie de la langue*. Mémoire de DEA en Sciences Cognitives, Institut National Polytechnique de Grenoble, 32 p.
- Martin, P. (1982) Phonetic realisation of prosodic contours in French. *Speech Communication*, 1, pp. 282-294.
- Mehler, J., Dommergues, J.-Y., Frauenfelder, U. & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal, Learning and Verbal Behavior*, 20, pp. 298-305.
- Mehler, J., Dupoux, E. & Segui, J. (1990) Constraining models of lexical access: the onset of word recognition. In *Cognitive Models of Speech Processing*, G.T.M. Altmann (ed.), Cambridge, MA: MIT Press, pp. 236-261.
- Meunier, C., Frenck-Mestre, C., Lelekov-Boissard, T. & Le Besnerais, M. (2003) Production and perception of foreign vowels: does the density of the system play a role? *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain.
- Mooshammer, C., Perrier, P., Fuchs, S., Geng, C. & Pape, D. (2004) An EMMA and EPG study on token-to-token variability. *Arbeitsberichte Institut für Phonetik und digitale Sprachverarbeitung Universität Kiel (AIPUK)*, Vol. 36, pp. 46-63.
- Munhall, K., Löfqvist, A. & Kelso, J.A.S. (1986) Laryngeal compensation following sudden oral perturbation. *Journal of the Acoustical Society of America*, 80, Suppl. 1, S109.
- Munhall, K. & Löfqvist, A. (1992) Gestural aggregation in speech: Laryngeal gestures. *Journal of Phonetics*, 20, pp. 111-126.
- Nelson, W.L. (1983) Physical principles for economies of skilled movements. *Biological Cybernetics*, 46, pp. 135-147.

- Nguyen, N. & Marchal, A. (1993) Assessment of an alternating magnetic fields device for tracking articulatory movement in speech production. *Journal of the Acoustical Society of America*, 94, pp. 1152-1155.
- Nguyen, N., Hoole, P. & Marchal, A. (1994) Regenerating the spectral shape of [s] and [S] from a limited set of articulatory parameters. *Journal of the Acoustical Society of America*, 96, pp. 33-39.
- Nguyen, N., Marchal, A. & Content, A. (1996) Modelling tongue palate contact patterns in the production of lingual consonants. *Journal of Phonetics*, 24, pp. 77-97.
- Nguyen, N. & Fagyal, Z. (2007) Acoustic aspects of vowel harmony in French. *Journal of Phonetics* (In Press).
- Öhman, S. (1966) Coarticulation in VCV utterance: Spectrographic measurements. *Journal of the Acoustical Society of America*, 39, pp. 151-168.
- Öhman, S. (1967) Numerical model of coarticulation. *Journal of the Acoustical Society of America*, 41, pp. 310-320.
- Ostry, D.J., Gribble, P.L. & Gracco, V.L. (1996) Coarticulation of jaw movements in speech production: Is context sensitivity in speech kinematics centrally planned? *Journal of Neuroscience*, 16(4), pp. 1570-1579.
- Pasdeloup, V. (1990) *Modèle de règles rythmiques du français appliqué à la synthèse de la parole*. Thèse de 3^e cycle en Phonétique, Université de Provence, 386 p.
- Payan, Y. & Perrier, P. (1997) Synthesis of V-V Sequences with a 2D biomechanical tongue model controlled by the Equilibrium Point Hypothesis. *Speech Communication*, 22(2/3), pp. 185-205.
- Perkell, J.S. (1969) *Physiology of speech production: Results and implications of a quantitative cineradiographic study*. MIT Press, Cambridge, Mass: MIT Press.
- Perkell, J.S. (1990) Testing theories of speech production: implication of some detailed analyses of variable articulatory data. In W.J. Hardcastle & A. Marchal (eds), *Speech Production and Speech Modelling*, The Netherlands: Kluwer Academic Publishers, pp. 263-288.
- Perkell, J.S., Guenther, F.H., Lane, H., Matthies, M.L., Perrier, P., Vick, J., Wilhelms-Tricarico, R. & Zandipour, M. (2000) A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics*, 28, pp. 233-272.
- Perkell, J.S., Matthies, M.L., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J. & Guidé, P. (1997) Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. *Speech Communication*, 22, pp. 227-250.

- Perrier, P., Boë, L.-J. & Sock, R. (1992) Vocal Tract Area Function Estimation from Midsagittal Dimensions with CT Scans and a Vocal Tract Cast: Modelling the Transition with Two Sets of Coefficients. *Journal of Speech and Hearing Research*, 35, pp. 53-67.
- Perrier, P., Ostry, D. J. & Laboissière, R. (1996a) The equilibrium point hypothesis and its application to speech motor control. *Journal of Speech and Hearing Research*, 39(2), pp. 365-378.
- Perrier, P., Loevenbruck, H. & Payan, Y. (1996b) Control of tongue movements in speech: the equilibrium point hypothesis perspective. *Journal of Phonetics*, 24, pp. 53-75.
- Perrier, P., Perkell, J., Payan, Y., Zandipour, M., Guenther, F. & Khalighi, A. (2000) Degrees of freedom of tongue movements in speech may be constrained by biomechanics. In *Proceedings of the 6th International Conference on Spoken Language Processing, ICSLP'2000*, 2, Beijing, China, pp. 162-165.
- Perrier P., Payan Y., Zandipour M. & Perkell, J. (2003) Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. *Journal of the Acoustical Society of America*, 114(3), pp. 1582-1599.
- Perrier, P., Payan, Y. & Marret, R. (2004). Modéliser le physique pour comprendre le contrôle : le cas de l'anticipation en production de parole. In R. Sock & B. Vaxelaire (éds) *L'anticipation à l'horizon du présent*, Sprimont, Belgique : Pierre Margala éd., pp. 159-177.
- Perrier, P., Ma, L. & Payan, Y. (2005a) Modeling the production of VCV sequences via the inversion of a biomechanical model of the tongue. *Proceedings of Interspeech 2005*, Lisbon, Portugal, pp. 1041-1044.
- Perrier, P. (2005b) Control and representation in speech production. *ZAS papers in Linguistics* 40, pp. 109-132.
- Perrier, P. (2006) About Speech Motor Control Complexity. In J. Harrington & M. Tabain (eds), *Speech Production: Models, Phonetics Processes, and Techniques*, Psychology Press, pp. 13-25.
- Petursson, M. & Bothorel, A. (1973) Etude sur la coarticulation en français dans des logatomes du type VCV. *Travaux de l'Institut de Phonétique de Strasbourg*, 5, pp. 100-128.
- Poggio, T. & Girosi, F. (1989) *A theory of networks for approximation and learning*. Technical Report A. I. Memo, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.
- Recasens, D. (1984) Vowel-to-vowel coarticulation in catalan VCV sequences. *Journal of the Acoustical Society of America*, 76, pp. 1624-1635.

- Recasens, D. (1987) An Acoustic Analysis of Vowel-to-Consonant and Vowel-to-Vowel Coarticulatory Effects in Catalan and Spanish VCV Sequences. *Journal of Phonetics*, 15, pp. 299-312.
- Recasens, D. (2002) An EMA study of VCV coarticulatory direction. *Journal of the Acoustical Society of America*, 111, pp. 2828-2841.
- Rousselot, P.-J. (1901) *Principes de phonétique expérimentale*, vol.1, Paris : H. Welter.
- Rossi, M. (1979) Le français, langue sans accent ? In I. Fonagy & P. Léon (éds), *L'accent en français contemporain*, *Studia Phonetica* 15, Paris : Didier, pp. 13-51.
- Rubin, P., Baer, T. & Mermelstein, P. (1981) An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America*, 70, pp. 321-328.
- Rumelhart, D.E. & Norman, D.A. (1982) Simulating a skilled typist: A study of skilled cognitive-motor performance. *Cognitive Science*, 6, pp. 1-36.
- Rumelhart D.E, Hinton G.E. & Williams, R.J. (1986) Learning internal representations by error propagation. In D.E. Rumelhart and J.L. McClelland (eds), *Parallel Distributed Processing: Exploration in the microstructure of cognition*, vol. 1, Cambridge, MA: MIT Press, pp. 318-362.
- Saltzman, E.L. (1986) Task dynamic coordination of the speech articulators: A preliminary model. *Experimental Brain Research, Ser. 15*, pp. 129-144.
- Saltzman, E.L. & Munhall, K.G. (1989) A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1, pp. 333-382.
- Savariaux, C., Perrier, P. & Orliaguet, J.-P. (1995) Compensation strategies for the perturbation of the rounded vowel [u] using a lip-tube: a study of the control space in perturbation. *Journal of the Acoustical Society of America*, 98, pp. 2428-2442.
- Savariaux, C., Perrier, P. & Orliaguet, J.-P. & Schwartz J.-L. (1999) Compensation for the perturbation of French [u] using a lip tube: II. Perceptual analysis. *Journal of the Acoustical Society of America*, 106 (1), pp. 381-393.
- Schwartz, J.-L., Beautemps, D., Abry, C. & Escudier, P. (1993) Inter-individual and cross-linguistic strategies for the production of the [i] vs [y] contrast. *Journal of Phonetics*, 21, pp. 411-425.
- Schwartz, J.-L., Abry, C., Boe, L.-J., & Cathiard, M.-A. (2002) Phonology in a theory of perception-for-action-control. In J. Durand, & B. Laks (eds), *Phonetics, phonology and cognition*, Oxford: Oxford University Press, pp. 255-280.
- Siegel, G. & Pick, H. (1974) Auditory feedback in the regulation of voice. *Journal of the Acoustical Society of America*, 65, pp. 1618-1624.

- Sock, R. (1998) *Organisation temporelle en production de la parole. Emergence de catégories sensori-motrices phonétiques*. Doctorat d'état de Sciences du Langage, Université Stendhal - Grenoble III, 479 p.
- Sock, R., Hecker, V. & Cathiard, M.-A. (1999) The perceptual effects of anticipatory labial activity in French. *14th International Congress of Phonetic Sciences, Vol 3*, San Francisco, pp. 2057-2060.
- Sock, R., Vaxelaire, B. & Hecker, V. (2002) What relationship between protrusion anticipation and auditory perception ? *7th International Conference on Spoken Language Processing*, Denver, Colorado, USA, pp. 1677-1680.
- Stevens, K.N. (1972) The Quantal Nature of Speech : Evidence from Articulatory - Acoustic Data. In E.E. David & P.B. Denes (eds), *Human Communication: A Unified View*, McGraw-Hill, New York, pp. 51-66.
- Stevens, K.N. & Blumstein, S.E. (1978) Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 64, pp. 1358-1368.
- Stevens, K.N. (1989) On the quantal nature of speech. *Journal of Phonetics*, 17, pp. 3-45.
- Stevens, K.N. (1998) *Acoustic phonetics*. MIT Press, Cambridge.
- Straka, G. (1950) *Système des voyelles du français moderne*. Bulletin de la faculté des Lettres de Strasbourg.
- Sussman, H.M., Bessell, N., Dalston, E. & Majors, T. (1997) An investigation of stop place of articulation as a function of syllable position: A locus equation perspective. *Journal of the Acoustical Society of America*, 101, pp. 2826-2838.
- Turvey, M.T. (1977) Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (eds), *Perceiving acting and knowing: Toward an ecological psychology*, Hillsdale, NJ: LEA.
- Ushijima, T. & Hirose, H. (1974) Electromyographic study of the velum during speech. *Journal of Phonetics* 2, pp. 315-326.
- Vaxelaire, B., Sock, R., Bonnot, J.-F. & Keller, D. (1999) Anticipatory labial activity in the production of French rounded vowels. X-ray and acoustic data. *14th International Congress of Phonetic Sciences, vol.1*, San Francisco, USA, pp. 53-56.
- Vaxelaire, B., Sock, R., Roy, J.-P., Ascì, A. & Hecker, V. (2003) Audible and inaudible anticipatory gestures in French. *15th International Congress of Phonetic Sciences*, pp. 447-450, Barcelona.
- Vatikiotis-Bateson, E., Tiede, M., Wada, Y., Gracco, V. & Kawato, M. (1994) Phoneme extraction using via point estimation of real speech. In *Proceedings of the International Conference on Spoken Language Processing*, 2, Yokohama, pp. 631-634.

- Whalen, D. H. (1990) Coarticulation is largely planned. *Journal of Phonetics*, 18, pp. 3-35.
- Wood, S.A.J. (1991) X-ray data on the temporal coordination of speech gesture. *Journal of Phonetics*, vol. 19, pp. 281-292.
- Wood, S.A.J. (1996) Assimilation or coarticulation? Evidence from the coordination of the tongue gestures for the palatalization of Bulgarian alveolar stops. *Journal of Phonetics*, vol. 24, pp. 139-164.
- Wood, S.A.J. (1997) A cinefluographic study of temporal organization of articulator gestures: Examples from Greenlandic. *Speech Communication*, vol 22, pp. 207-225.
- Wu, Z. (1989) *Outline of Experimental Phonetics*. High Education Press, China.
- Wu, Z. & Sun, G. (1990) Acoustic coarticulatory patterns of voiceless fricatives in CVCV in Standard Chinese. *RPR-IL (CASS: Chinese Academy of Social Sciences)*, pp. 1-19.
- Wu, Z. & Sun, G. (1991) A study of coarticulation of unaspirated stops in CVCV contextes in Standard Chinese. *Proceedings of the 12th International Congress of Phonetic Sciences*, Vol. 3, Aix-en-Provence, France, pp. 374-377.
- Wu, Z. (2000) From Traditional Chinese Phonology to Modern Speech Processing: Realization of Tone and Intonation in Standard Chinese. *International Conference on Spoken Language Processing*, 1, Beijing, China, B1-B12.
- Xu, Y. (1997) Contextual Tonal Variation in Mandarin. *Journal of Phonetics*, 25(1), pp. 61-83.
- Xu, Y. & Liu, F. (2006). Tonal alignment, syllable structure and coarticulation: Toward an integrated model. *Italian Journal of Linguistics*, 18, pp. 125-159.
- Zerling, J.-P. (1979) *Articulation et coarticulation dans des groupes occlusive-voyelle en français*. Thèse de Doctorat en Phonétique, Université de Nancy II, 344 p.
- Zerling, J.-P. (1991) Articulation, coarticulation et contraintes: quelques points de vue d'auteurs. *Travaux de l'Institut de Phonétique de Strasbourg*, 22, pp. 87-104.
- Zerling, J.-P. (1993) Les contraintes en production de la parole. *Travaux de l'Institut de Phonétique de Strasbourg*, 23, pp. 177-181.
- Zheng, Y.L. & Bao, H.Q. (2003) On the coarticulation of /-N1C2/ sequences in Standard Chinese. *Proceedings of the 6th National Conference on Modern Phonetics*, Tianjin Normal University Press, China, pp. 37-43.

La coarticulation en français et en chinois : Étude expérimentale et modélisation

Résumé: Comme d'autres mouvements humains, les gestes de la parole sont vraisemblablement planifiés selon des stratégies optimales. Cependant, il est probable que la planification de ces gestes soit aussi contrainte par des critères de nature linguistique. De telles contraintes pourraient porter notamment sur la structure phonologique et la longueur des séquences prises en compte dans la planification. Le but de cette thèse a été d'approfondir cette hypothèse à partir de l'analyse de données expérimentales et de simulations avec un modèle.

Pour cela, la coarticulation anticipante a été étudiée expérimentalement dans des séquences Voyelle₁-Consonne-Voyelle₂ (V₁CV₂) pour deux langues différentes, le français et le chinois, à partir de données articulatoires. Les résultats de nos analyses montrent que les locuteurs français et chinois ont adopté des stratégies de planification différentes : en français l'anticipation s'étend au delà des frontières de la syllabe CV, alors qu'en chinois ce n'est pas le cas. Pour valider cette interprétation sur le rôle de la syllabe, un modèle de contrôle moteur a été élaboré exploitant une stratégie de planification qui minimise un critère combinant effort articulatoire et contraintes perceptives, et il a été testé sur un modèle biomécanique de la langue. Différentes hypothèses sur la structure phonologique de la séquence planifiée ainsi que sur l'exécution du mouvement ont été testées dans ce cadre. Les simulations ont été comparées aux données. Nos résultats soutiennent l'hypothèse selon laquelle les influences de la syllabe seraient plus fortes en chinois qu'en français.

Mots clés: coarticulation, anticipation, planification, articulation, syllabe, phonologie, contrôle moteur, commande motrice, optimisation.

Coarticulation in French and in Chinese: Experimental study and modelling

Abstract: As other skilled human movements, speech gestures are likely to be planned according to optimal motor control strategies. However, it can be assumed that speech production planning is, in addition, constrained by linguistic factors. Such constraints could determine the phonological structure and the length of the speech sequences on which optimal planning applies. The purpose of this thesis work was to study this hypothesis using experimental data and speech production models.

Anticipatory coarticulation within Vowel₁-Consonant-Vowel₂ (V₁CV₂) sequences was experimentally studied for two different languages, French and Mandarin Chinese, using articulatory data. Results have shown that, for French speakers, V₂ influences the articulation of the whole sequence V₁CV₂, while its influence is strictly limited to the CV₂ syllable for Mandarin speakers. This suggests that speech planning in French extends beyond the limits of the syllable, contrary to Mandarin. To quantitatively validate this interpretation, a control model of planning was developed and applied to a biomechanical tongue model. Planning consists in minimizing a criterion combining articulatory effort and perceptual constraint. Different hypotheses concerning the phonological structure of the planned sequence and the execution of movement were tested. Simulated coarticulation patterns were compared to the experimental ones. In sum, experimental and modeling findings support the hypothesis of a stronger influence of the syllable in Chinese than in French.

Keywords: coarticulation, anticipation, motor planning, articulation, syllable, phonology, motor control, command, optimisation.

LPL, CNRS UMR 6057, 29, Avenue Robert Schuman, 13621 Aix en Provence.

GIPSA-Lab, DPC, CNRS UMR 5216, 46, Avenue Félix Viallet, 38031 Grenoble Cedex.